12-2017

The Effects of Repeated Lineups and Delay on Eyewitness Identification

Wenbo Lin
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological & Brain Sciences

The Effects of Repeated Lineups and Delay on Eyewitness Identification

by

Wenbo Lin

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Masters of Arts

December 2017
St. Louis, Missouri

Table of Contents

List of Figures iv

List of Tables v

Acknowledgments vi

Abstract vii

Chapter 1: Introduction 1

1.1 The Effects of Repeated Exposure 5

1.1.1 Identification Accuracy 5

1.1.2 The Confidence-Accuracy Relationship 7

1.2 The Effects of Delay 9

1.2.1 Identification Accuracy 9

1.2.2 The Confidence-Accuracy Relationship 11

1.3 First-Year Project Results 12

1.4 Response Times as an Indicator of Accuracy 15

1.5 Overview of Experiments 16

Chapter 2: Experiment 1 17

2.1 Method 20

2.1.1 Participants 20

2.1.2 Design 20

2.1.3 Materials 21

2.1.4 Procedures 22

2.2 Results 24

2.2.1 Frequency of Response Types 24

2.2.2 The Effects of Repeated Lineups and Delay on Choosing 24

2.2.3 The Effects of Repeated Lineups and Delay on Accuracy 27

2.2.4 Identification Decisions and Confidence Ratings 28

2.2.5 Confidence-Accuracy Relationship 34

2.2.6 Confidence Ratings and Response Times as Indicators of Identification Accuracy 38

2.2.7 Consistent and Inconsistent Decisions across Lineup 1 and 2 40

2.2.8 Lineup 1 and 2 Decision Combinations and Their Confidence Ratings 42

2.2.9 Confidence and Decision Consistency 45

2.3 Discussion 47
Chapter 3: Experiment 2 52
3.1 Method 54
3.1.1 Participants 54
3.1.2 Design 54
3.1.3 Materials 57
3.1.4 Procedures 57
3.2 Results 60
3.2.1 Frequency of Response Types 60
3.2.2 The Effects of Repeated Lineups and Delay on Choosing 60
3.2.3 The Effects of Repeated Lineups and Delay on Accuracy 64
3.2.4 Identification Decisions and Average Confidence 65
3.2.6 Confidence Ratings and Response Times as Indicators of Identification Accuracy 74
3.2.7 Consistent and Inconsistent Decisions across Lineup 1 and 2 77
3.2.8 Lineup 1 and 2 Decision Combinations and Their Confidence Ratings 79
3.2.9 Confidence and Decision Consistency 84
3.3 Discussion 84
Chapter 4: General Discussion 90
4.1 Choosing 90
4.2 Accuracy 91
4.3 Decision Consistency 92
4.4 Confidence-Accuracy Relationship 93
4.5 Confidence Judgments versus Response Times 94
4.6 Limitations 96
4.7 Implications 97
References 99
Appendix A 103
Appendix B 113

List of Figures

Figure 1: Eyewitness Identification Decisions ……………………………………………...….3

Figure 2: First-Year Project Design …………………………………………………...…13

Figure 3: Experiment 1 Design …. …..………………………………………………...…18

Figure 4: Experiment 1: Average Confidence and Target-Present Choosing Status………..…31

Figure 5: Experiment 1: Average Confidence and ID Responses…....……...……………......31

Figure 6: Experiment 1: Average Confidence and Target-Absent Choosing Status………..…33

Figure 7: Experiment 1: Chooser Calibration Plots…….……………………….……..…37

Figure 8: Experiment 1: Nonchooser Calibration Plots…….……………..………………...…37

Figure 9: Experiment 2 Design …. …..………………………………………………...…58

Figure 10: Experiment 2: Average Confidence and Target-Present Choosing Status……...…66

Figure 11: Experiment 2: Average Confidence and ID Responses…....……...…..……….......68

Figure 12: Experiment 2: Average Confidence and Target-Absent Choosing Status……...…70

Figure 13: Experiment 2: Chooser Calibration Plots…….……………………………....…72

Figure 14: Experiment 2: Nonchooser Calibration Plots…….………………..…………....…72

Figure A1: Experiment 1: Average RT and Target-Present Choosing Status……………......104

Figure A2: Experiment 1: Average RT and ID Responses…....……...……………..............104

Figure A3: Experiment 1: Average RT and Target-Absent Choosing Status……………......107

Figure A4: Experiment 2: Average RT and Target-Present Choosing Status…………….....109

Figure A5: Experiment 2: Average RT and ID Responses…....……...……………..............111

Figure A6: Experiment 2: Average RT and Target-Absent Choosing Status……………......111

List of Tables

Table 1: Experiment 1: Identification Responses in Target-Present Lineups…..…………...…25

Table 2: Experiment 1: Identification Responses in Target-Absent Lineups…..…………...…25

Table 3: Experiment 1: Decision Shift……………………………….…..…………...…29

Table 4: Experiment 1: Confidence Bins……………………………….…..…………...…35

Table 5: Experiment 1: Accuracy of Responses with a High Confidence criterion………..…39

Table 6: Experiment 1: Accuracy of Responses with a Fast RT criterion…………………….39

Table 7: Experiment 1: Accuracy of Responses with High Confidence and Fast RT…..…….39

Table 8: Experiment 1: Proportion of Same and Different Decisions……………………...…41

Table 9: Experiment 1: Confidence Ratings of Same and Different Decisions…………...…41

Table 10: Experiment 1: Decision Combinations and Confidence Ratings…..…………...…43

Table 11: Experiment 1: Consistent Responses by Confidence Level……….…………...…46

Table 12: Experiment 2: Identification Responses in Target-Present Lineups…..…………...61

Table 13: Experiment 2: Identification Responses in Target-Absent Lineups…..……………62

Table 14: Experiment 2: Confidence Bins……………………………….…..…………....73

Table 15: Experiment 2: Accuracy of Responses with a High Confidence criterion…………75

Table 16: Experiment 2: Accuracy of Responses with a Fast RT criterion…………………..76

Table 17: Experiment 2: Accuracy of Responses with High Confidence and Fast RT…..…..76

Table 18: Experiment 2: Proportion of Same and Different Decisions……………………....78

Table 19: Experiment 2: Confidence Ratings of Same and Different Decisions…………....78

Table 20: Experiment 2: Target-Present Decision Combinations and Confidence Ratings….80

Table 21: Experiment 2: Target-Present Decision Combinations and Confidence Ratings….81

Table 22: Experiment 2: Consistent Responses by Confidence Level……….………......83

Acknowledgments

I would like to thank my graduate advisor, Roddy Roediger, for his support and mentorship. I would also like to thank the members of my defense committee, Dave Balota and Andrew Butler, for their insight and advice. I am grateful to Mike Strube for his help with the statistical analyses. I also want to recognize members of the Memory Lab, Oyku Uner and Eylul Tekin, for their support.

Funding was provided by Washington University Research Fund.

Wenbo Lin

Washington University in St. Louis

December 2017

Abstract

The Effects of Repeated Lineups and Delay on Eyewitness Identification

by

Wenbo Lin

Master of Arts in Psychological & Brain Sciences

Washington University in St. Louis, 2018

Professor Henry L. Roediger, III

Prior eyewitness research has examined the effects of repeated identification procedures and delays on eyewitness identification, but these studies have either confounded these two factors or studied them in isolation. Experiment 1 attempted to disentangle these factors through systematic manipulations of the number of repeated lineups and the length of delay between the original event and the first lineup. Experiment 2 examined whether the length of delay between two lineups (Lineups 1 and 2) affects the subsequent lineup identification decisions. We found that people were more inclined to choose when a lineup was repeated. A longer delay between the crime and the initial identification decreased the tendency to choose for target-present lineups but not for target-absent lineups. Regardless of delay conditions, both correct and incorrect decisions in Lineup 1 were often repeated in Lineup 2. Compared to response times, confidence was a more reliable indicator of identification accuracy. More importantly, in both experiments, the confidence-accuracy relationship remained intact despite the effects of repeated lineups and delay.

Chapter 1: Introduction

Eyewitness identification is one of the most compelling types of evidence in criminal trials, but it is also the leading cause of wrongful convictions (www.innocenceproject.org). Erroneous identification played a role in over 70 percent of DNA exoneration cases. Such unsettling numbers of wrongful convictions warrant both increased research effort and procedural reform. Over the years, researchers have evaluated the utility of various identification procedures: mug-shot (a single-suspect photo), show-up (a live presentation of a single suspect), simultaneous lineup (all the suspects are presented at once) and sequential lineup (each suspect is presented one at a time). An equally important issue, however, is how typical identification practices might affect the reliability of eyewitness identification. This includes repeated exposure (e.g. seeing a face multiple times in repeated identification procedures) and retention interval (i.e. the delay between the crime and first identification). Do these factors reduce the effectiveness of these identification procedures and the reliability of eyewitness identification?

Eyewitness misidentification cases involving repeated exposure are not rare. To date, there have been several known eyewitness misidentification cases involving multiple identifications (Garrett, 2011). The widely cited Ronald Cotton case is an example of how repeated identifications led to disastrous consequences. In July 1984, Bobby Poole broke into Jennifer Thompson's apartment and sexually assaulted her. When Thompson was asked to identify her assailant, she misidentified Ronald Cotton at both the photo lineup and the subsequent live lineup. Poole was not in either lineups. Although Thompson also identified Cotton at court and said that she was "absolutely sure", she was neither confident at the photo lineup nor at the live lineup. She became confident over time because she received confirmatory post-identification feedback from the police, a practice now known to inflate eyewitness confidence (Bradfield, Wells, & Olson, 2002; Quinlivan, Neuschatz, Douglass, Wells, & Wetmore, 2012; Semmler, Brewer, & Wells, 2004; Wells & Bradfield, 1998). Thompson did not come face to face with Poole until years later. When she did, she could not recognize Poole. Her response was: "I have never seen him in my life. I have no idea who he is." Without the help of DNA testing, Cotton would still be in prison today for a crime he did not commit. The Ronald Cotton case was highly publicized, but it is not a unique case of eyewitness misidentification involving repeated identifications. David Lee Wiggins was convicted for the rape of a 14-year-old girl in Fort Worth, Texas. The victim picked him at the photo and live lineups. Although his fingerprint did not match those found at the crime scene, he was sentenced to life in prison. Fortunately, DNA testing was able to exonerate him. These are just two known misidentification cases where DNA evidence was available. The exact statistic of wrongful convictions involving repeated identifications is unknown.

Repeated identification procedures are relatively routine (Behrman & Davey, 2001; Steblay, 2011), but they may complicate the identification process. Witnesses may initially see a mugshot (or a show-up) followed by a lineup, or they may see consecutive lineups (e.g. a photo lineup followed by a live lineup). In a typical identification, there are at least four outcomes (see Figure 1). Unlike other identification formats (e.g. mugshots and show-up), lineups allow eyewitnesses to choose a known innocent filler (or a foil). Because the police know that these fillers are innocent, there is no legal implication for filler identifications. For simplicity's sake, the following example will focus only on the four shared outcomes in eyewitness identification procedures. For example, in a target-present lineup (i.e. the guilty suspect is the lineup), witnesses either pick the guilty suspect (a hit or correct identification) or reject the lineup (a miss or incorrect rejection). In a target-absent lineup (i.e. the guilty suspect is not in the lineup),

Witness Decision

True Suspect Status	**Choose**	**Reject**	Suspect Presence
Guilty	Hit or Correct Identification	Miss or Incorrect Rejection	**Present**
Innocent	False Alarm or Incorrect Identification	Correct Rejection	**Absent**

Figure 1. The four common outcomes in eyewitness identification.

witnesses either pick an innocent suspect/filler (a false alarm) or reject the lineup (a correct rejection). With repeated identification procedures, witnesses can make different decisions across identifications.

Ideally, witnesses pick the guilty suspect in the first identification and again in the second identification. In this case, the second identification confirms the initial identification. This assumes that the second identification is independent of the first identification, which most likely is not true. Although the initial identification may tap into the witness's recognition memory for the original event (the crime), the second identification may not because of the contamination of memory by the first lineup. For instance, witnesses may not pick anyone in the initial lineup but choose a particular member in a subsequent lineup simply because this member is the only face to appear in both lineups. When a particular face is repeated across lineups, witnesses may unintentionally misinterpret the police's intention and believe that particular member is the perpetrator. Alternatively, this scenario may occur because witnesses confuse the source of their familiarity (Johnson, Hashtroudi, & Lindsay, 1993; Lindsay & Johnson, 1989). Witnesses may misattribute their familiarity for their selection in the subsequent lineup to a false memory for the perpetrator of the crime (misplaced familiarity). Another scenario occurs when witnesses select an innocent suspect in the initial lineup, and make the same decision in a subsequent lineup in order to remain consistent with their initial decision (a consistency effect). Eyewitnesses may perceive inconsistent decisions across lineups as an indication of unreliability; therefore, they feel compelled to stick with their initial decision regardless of whether it was a correct or incorrect identification. In short, repeated identification procedures are likely to elevate the risk of identification errors.

1.1 The Effects of Repeated Exposure

1.1.1 Identification Accuracy

Laboratory research has consistently shown that repeated identification procedures increase the risk of identification errors. Brigham and Cairns (1988) examined how prior exposure to mugshots affected the subsequent lineup accuracy. Both the control group and the experimental group watched a video of a staged crime. They also saw a target-present lineup after a 2-day delay. However, only the experimental group was exposed to intervening mugshots of innocent suspects prior to the lineup; the control group only saw the lineup after a 2-day delay. The mugshot photographs never included the suspect's photo. After a 2-day delay, both groups saw a six-person target-present lineup that included both the suspect and their mugshot choice (if participants had picked someone in the mugshot phase). Compared to the control group, the experimental group had a significantly lower hit rate when they had previously identified someone from the mugshots (.27 vs. .69) and even when they had not previously identified someone from the mugshots (.41 vs. .69). The experimental group was also more likely to false alarm relative to the control group (.29 vs .08). Thus, prior exposure to mugshots can decrease accuracy and increase false alarms in a subsequent lineup. This finding was confirmed in a a meta-analysis of similar research conducted by Deffenbacher, Bornstein, and Penrod (2006). The prior exposure to mugshots decreased the accuracy in a subsequent lineup when the same face from the mugshots was presented again at the lineup, a phenomenon now known as the mugshot exposure effect. That is, witnesses may pick an innocent person in subsequent lineups because they initially saw the same person in the mugshots. Prior exposure to mugshots has been shown to not only significantly decrease correct identifications but also increase false alarms.

Prior exposure to a show-up or a lineup produces similar effects. Godfrey and Clark (2010) found exposure to a show-up a week prior to a lineup increased both guilty suspect and innocent suspect identification rate, although the latter was not statistically reliable. Steblay, Tix, and Benson (2013) reported similar findings with two simultaneous or two sequential lineups. However, Steblay et al. used the same guilty (or innocent) suspect photo across the repeated lineups (i.e. Lineup 1 and Lineup 2), but they also used different sets of fillers for Lineup 1 and Lineup 2. Therefore, the guilty (or innocent) suspect photo was repeated across both lineups. Nevertheless, false identifications at the first lineup were more likely to be carried over to the second lineup than to be corrected. This pattern was observed in both simultaneous and sequential lineups, although it was less severe for sequential lineups due to the lower number of errors at the first lineup, which in turn limited the number of carryover errors. Similarly, Hinz and Pezdek (2001) found mere exposure to the innocent suspect in a prior target-absent lineup increased the false alarm rate in a subsequent target-absent lineup. Furthermore, the hit rate was lower even when the actual suspect was presented alongside the innocent suspect (.42) than when suspect was presented alone (.76). In other words, the repeated innocent suspect in the subsequent lineup lowered the witness's ability to identify the guilty suspect.

The undesirable consequences associated with repeated identification procedures raise a theoretical question concerning how a prior exposure or identification affects a subsequent identification. Researchers have considered two possibilities. First, witnesses may make the same identification in both the first and second identification because they are committed to their initial decision. Valentine, Davis, Memon, and Roberts (2012) reported that a majority of choosers (88%) in the show-up also selected the same suspect in the subsequent lineup, regardless of whether the identification was correct or incorrect. Likewise, witnesses who initially rejected the lineup may reject a subsequent lineup. Godfrey and Clark (2010) collectively referred to these repeated decisions as a consistency effect.

Second, witnesses might misperceive their sense of familiarity for an innocent suspect after seeing that innocent suspect in an earlier identification (Memon, Hope, Bartlett, & Bull, 2002). In other words, witnesses may pick an innocent suspect because they misattribute their familiarity with the innocent suspect for their memory of the true suspect, thereby committing a source-monitoring error (Johnson et al., 1993). According to Godfrey and Clark (2010), a signature pattern of misplaced familiarity is a *no-to-suspect* shift, which occurs when witnesses shift from being a non-chooser at the initial identification to being a chooser of a filler face in a subsequent identification. In sum, the consistency effect and misplaced familiarity can operate in tandem and negatively affect eyewitness identification decisions.

1.1.2 The Confidence-Accuracy Relationship

Does repeated exposure affect eyewitness confidence? Eyewitness confidence has been widely used and endorsed as an indicator of accuracy by U.S. Supreme Court, police, and jurors (Brewer & Burke, 2002; Deffenbacher & Loftus, 1982; Potter & Brewer, 1999). The confidence-accuracy (CA) relationship in eyewitness identification has been studied extensively (Bothwell, Deffenbacher, & Brigham, 1987; Juslin, Olsson, & Winman, 1996; Sporer, Penrod, Cutler, & Read, 1995). Accurate identification and confidence generally produce a positive CA relationship, whereas correct rejection and confidence generally produce a weak CA relationship (Brewer & Wells, 2006; Palmer, Brewer, Weber, & Nagesh, 2013; Sauer, Brewer, Zweck, & Weber, 2010; Sporer et al., 1995; Tekin & Roediger, 2017; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015; Wixted, Read, & Lindsay, 2016; Wixted & Wells, 2017). If repeated identification procedures introduce unintended eyewitness errors, then the CA relationship is also jeopardized. Imagine the following scenario: a witness identifies a suspect with low confidence at the first identification but become highly confident in court. Such a scenario is neither impossible nor unimaginable. During the criminal investigation of the Ronald Cotton case, Thompson showed signs of uncertainty prior to her "absolutely sure that Ronald Junior Cotton is the man." statement in court. When Thompson was initially presented with a lineup, she did not immediately identify Cotton. After a long decision process, she eventually settled on Cotton. She then asked one of the detectives, “Did I do OK?” The detective replied, “You did great, Ms. Thompson.” Likewise, Thompson was not confident at the live lineup. She stated that Cotton “looks the most like him.” Rather than asking how certain she was, the detective asked if she was certain and added, “We thought that might be the guy. It’s the same person you picked from the photos.” The confirmatory post-identification feedback in both identifications contaminated Thompson’s confidence in her memory. These deleterious effects of confirmatory feedback have been replicated in several laboratory studies (Bradfield & Mcquiston, 2004; Wells & Bradfield, 1998; Wells, Olson, & Charman, 2003). In short, Thompson was certainly not “absolutely sure” at her earlier identifications but became so at court. If confidence judgments made in subsequent identifications are less diagnostic, then greater emphasis should be placed on the confidence judgment made in the initial identification. In fact, Wixted and colleagues (2015) reported that initial eyewitness confidence reliably and strongly predicts identification accuracy. Since repeated identification procedures are prevalent in criminal investigation, more extensive research is needed to examine the potential effects of repeated exposures on eyewitness identification.

Although very few studies have examined the effects of repeated exposure on the CA relationship, two studies do offer some insights. In Steblay et al. (2013), choosers were equally confident in both their first and second lineup identification regardless of whether they had selected the guilty suspect or an innocent suspect. Although the repeated exposure did not inflate confidence, the confidence was maintained for both correct and incorrect identifications. More telling results were observed for repeated identifications involving a show-up followed by a lineup (Godfrey & Clark, 2010). Consistent witnesses who identified the suspect in the show-up and again in the lineup were more confident in their lineup identification than their show-up identification. Compare to these consistent witnesses, witnesses who made a nonidentification in the show-up but identified the suspect in the subsequent lineup were less confident in their lineup identifications. In other words, confidence in the subsequent lineup depended on whether that witnesses consistently identified the suspect in both the show-up and lineup, or if they had committed a *no-to-suspect* shift (i.e. witnesses who initially reject but later identify someone in a subsequent identification). A similar pattern was also observed for innocent suspects; however, there were too few witnesses who made innocent suspect identifications to provide a reliable conclusion. Based on these available findings, there are some hints that repeated exposure jeopardizes the CA relationship.

1.2 The Effects of Delay

1.2.1 Identification Accuracy

In contrast to the more controllable repeated identification procedure, the retention interval between the time of the crime and the first lineup is an uncontrollable factor outside the lab. Depending on how soon the police are able to construct a lineup, this time interval will surely vary. Based on past research, it is reasonable to predict that identification performance decreases as retention interval increases, just as in all memory research. The eyewitness identification literature on the effects of delay, however, remains mixed.

Cutler, Penrod, O'Rourke, and Martens (1986) found a greater decrease in correct identifications after 28 days than after 7 days. However, this delay effect was eliminated when participants reread their own description of the suspect and the crime just prior to making the identification. In a more recent study by Sauer, Brewer, Zweck, and Weber (2010), participants either made an identification immediately or after a delay depending on their assigned condition. Although the retention interval of the delayed condition varied from 20 to 50 days with a mean of 23 days, it was collectively analyzed as a single delayed condition. Their results indicated that identification accuracy was greater in the immediate condition than in the delayed condition. This greater accuracy in the immediate condition was reflected in both the greater number of correct identifications and correct rejections. Therefore, the increase in retention interval had a predictable negative effect on accuracy. Egan and collegues (1977) manipulated the retention interval with a 2-day, 21-day, or 56-day delay, and observed an increase in false alarms but no decrease in correct identifications as a function of retention interval. Thus, the decrease in accuracy was driven by the increase in false alarms in this study.

An extensive study conducted by Shepherd (1983) found no significant decrease in correct identifications from the 1-week (65%) to the 1-month (55%) to the 3-month (50%) retention interval. A significant decrease was not observed until 11-months (10%). There was no significant increase in false alarms as the retention interval increased. Valentine, Pickering, and Darling (2003) obtained similar results. Suspect identification was at 65% when the identification occurred within a week. No significant decrease in suspect identification was observed when the identification occurred between 1-month and 6-months, or even after 6-months. Despite of the increasing delay, the suspect identification rate fluctuated in the range of 34-46% with no hint of a reliable decrease as a function of retention interval. Similar to

Shepherd (1983), Valentine and colleagues did not find an increase in foil identifications (i.e. known innocent fillers in a lineup) across the retention intervals.

These inconsistent findings could be due to differences in materials, procedures, experimental power, or sample characteristics. Nevertheless, meta-analyses have found negative effects of retention interval on identification performance (Deffenbacher, Bornstein, Mcgorty, & Penrod, 2008; Shapiro & Penrod, 1986). Perhaps the troubling yet unresolved question is how much does retention interval affect the eyewitness identification process? Deffenbacher et al. (2008) applied Wickelgren's power-exponential theory to fit forgetting functions for varying retention interval ranges and concluded that it is possible to estimate both the percentage of initial memory strength remaining at any retention interval and the probability of correct identification of a specified lineup size at a particular retention interval. However, this curve-fitting was only applied to 11 forgetting functions from eight published studies. Moreover, since neither laboratory nor field experiments fully replicate all the complex variables in forensic situations (e.g., perpetrator distinctiveness, witness attention, or the level of stress), the applicability of these predictions may be limited at this time. Thus far, retention interval does appear to affect identification performance, but further research is required to determine, with greater precision, the magnitude with which retention interval affects identification.

1.2.2 The Confidence-Accuracy Relationship

Perhaps more importantly, how does retention interval affect high confidence identification? A few studies have specifically examined the effects of retention interval on the CA relationship with the use of calibration curves. In Sauer et al. (2010), participants in the immediate condition had greater accuracy than the delayed condition and there was greater overconfidence in the delayed condition, yet retention interval did not significantly undermine the CA relationship even in the delayed condition. In other words, increased confidence continued to be associated with increased diagnosticity for choosers despite of the increase in delay. Palmer et al. (2013) reported similar findings. Calibration was better in the immediate condition than the delayed condition, but the overall positive confidence-accuracy association for choosers was still observed. This relationship remained true even under difficult conditions involving reduced exposure duration and divided attention manipulations. In a more recent study, Wixted, Read, and Lindsay (2016) reported that the confidence-accuracy relationship remained meaningful, and high confidence identifications continued to indicate high accuracy, even after a long retention interval of 9 months. Therefore, the confidence-accuracy relationship appears to be quite robust despite the effects of delay given that only a single test is given at varying length of delays.

1.3 First-Year Project Results

My first-year project examined both the effects of repeated exposure and retention interval on eyewitness identification. The experiment was a 3 (initial lineup delay: 1-week, 2-weeks, or 3-weeks) x 3 (lineup presentation: 1, 2, or 3 lineup tests) x 2 (lineup type: target-present versus target-absent) design. All manipulations were conducted within-subjects with four experimental sessions, spaced a week apart (see Figure 2). Participants were presented with twelve videos of twelve targets in the first session and made lineup identifications in session 2-4. Four lineups were presented in session 2 (Set A), eight in session 3 (Set A & B), and twelve in session 4 (Set A, B, & C). After session 2, each subsequent week's lineups consisted of the lineups from the previous week (e.g. Set A is repeated in session 3) plus four new lineups (e.g. Set B is presented for the first time in session 3). For each session, half of the lineups were target-present and half target-absent. The presentation order of these lineups was randomized

Week 1 Session 1	Week 2 Session 2	Week 3 Session 3	Week 4 Session 4
12 Videos	Lineup Set A	Lineup Set A	Lineup Set A
		Lineup Set B	Lineup Set B
			Lineup Set C

Figure 2. The first-year project design. Each lineup set consists of four lineups. The number of prior lineups was manipulated within subjects.

(i.e. they saw both the old and new lineups in random order). The presentation of a particular lineup on a particular test was counterbalanced across participants. For each lineup, participants made an identification and a confidence judgment.

Results showed that participants were more inclined to choose someone from a lineup when a particular lineup was repeated. However, they were less likely to choose someone from a lineup if they saw the lineup for the first time after a long delay. In other words, participants were more likely to choose someone from a lineup that was presented for the first time after a 1-week delay compared to a lineup that was presented for the first time after a 2-week delay. Similar patterns of results were also observed in guilty suspect and innocent suspect identifications. As the number of repeated exposures increased, the identification rate also increased. Likewise, the identification rate also decreased as the length of retention interval between the event (the videos) and the first lineup increased. Moreover, results suggested that consistent identification decisions (e.g. guilty suspect to guilty suspect or innocent suspect to innocent suspect identifications) across consecutive lineups showed signs of confidence inflation. However, a within-subject experiment like this is not reflective of the real world identification process because eyewitnesses usually only have to identify one suspect, not twelve. Furthermore, this multiple targets design may have introduced additional unintended effects. For example, participants might make their identification decision and confidence judgment for lineup B based on their memory of their decision and confidence made for lineup A even though the two lineups were completely unrelated (i.e. sequential judgment effects). Therefore, the patterns of results observed in my first-year project should be replicated in a between-subject design using only one target. Experiment 1 of the present study attempted to replicate these results through the use of a between-subject design. In addition to confidence judgments, response latency measures were also included as an additional post-dictive indicator of accuracy, for reasons discussed next.

1.4 Response Times as an Indicator of Accuracy

One potential alternative measure to confidence judgments in assessing witness accuracy is response latencies. Research has consistently found that correct identifications are made faster than incorrect identifications; however, like confidence judgment, response latency does not discriminate between correct and incorrect rejections (Brewer, Caon, Todd, & Weber, 2006; Dunning & Perretta, 2002; Smith, Lindsay, & Pryke, 2000; Weber & Brewer, 2006; Weber, Brewer, Wells, Semmler, & Keast, 2004). Furthermore, unlike a confidence judgment scale (e.g. 0 to 100%), there is no scale for response latency. In other words, there is no guideline as to what is fast or slow except in a relative sense (faster or slower than other judgements). Dunning and Perretta (2002) proposed a response time boundary rule of 10 to 12 seconds that best differentiate accurate (suspect IDs) from inaccurate identifications (innocent suspect or filler IDs). Later studies, however, obtained evidence inconsistent with this rule (Brewer et al., 2006; Weber et al., 2004). Brewer et al. (2006) found varying optimum time boundaries (i.e. the time interval that best discriminates correct from incorrect identifications) with different retention intervals and lineup sizes. Weber et al. (2004), however, found impressive identification accuracy when decisions were made with both high confidence (90-100% confidence) and fast response time (10-s time boundary). The combination of high confidence and fast response time was superior to the accuracy obtained by either confidence or response time alone. In sum, research seems to suggest that response latency measures are not a replacement for confidence judgments; however, they could potentially be a useful secondary post-dictive indicator of accuracy. Nonetheless, further research is required to examine robustness of response latencies under various manipulations before such measures can be implemented in real world forensic protocols.

1.5 Overview of Experiments

The present study consisted of two experiments that provide an extensive investigation into the joint effects of repeated exposure and retention interval on eyewitness identification involving consecutive lineups and their effects on witness decision-making behaviors, accuracy, confidence, and response latency. Prior research on repeated exposures generally involved a mugshot (or a show-up) followed by a lineup, but only a few studies have used repeated lineups. Furthermore, these repeated lineup studies often did not manipulate the length of delay between the crime and the first lineup or the length of delay between the two lineups. Therefore, these studies confounded repeated exposure with delay. Although there are more studies that examined delay effects than repeated lineup effects, they have focused on the length of delay between the crime and the first lineup. The delay between two lineups is often overlooked. In short, repeated lineups and delay differences are commonplace in the real world, but prior research either confounds them or examines them separately. This in turn limits both our understanding of their effects on eyewitness identification and our ability to generalize to real world forensic situations where both factors could co-occur.

The main objectives of the present study were to address the following questions: 1) What are the effects of repeated lineups and delays (e.g., the delay between the crime and the first lineup, and the delay between consecutive lineups) on witness decision behavior (e.g. the tendency to choose), accuracy, confidence, and response latency? 2) What are the effects of repeated lineups and delays on the confidence-accuracy relationship? 3) Is response time as reliable as confidence judgmentss in discriminating correct from incorrect identifications?

Chapter 2: Experiment 1

The purpose of Experiment 1 was to replicate the within-subjects design results from my first-year project. However, there were a few modifications. Experiment 1 used a 2 (lineup type: target-present or target-absent lineup) x 2 (initial delay: 10mins-3days or 3days-3days delay) x 2 (the number of lineups: Lineup 1 and Lineup 2) mixed design. The lineup type and delay were manipulated between-subjects, whereas the number of lineups was manipulated within-subjects. Participants only saw one person (or target) instead of twelve targets. Regardless of the initial delay between the event and Lineup1, all the participants saw their assigned lineup the same number of times (see Figure 3).

We hypothesized that choosing rate (i.e. selecting a suspect or foil) should increase as a function of repeated exposure due to increased familiarity from prior exposure. On the other hand, witnesses should be less inclined to select someone when there was a long delay between the event and the lineup because they would be less confident about their memory due to normal forgetting. Therefore, we expected choosing rate to decrease with an increase in retention interval. Based on these predictions, repeated exposures were expected to negatively affect the CA relationship. We hypothesized that retention interval should have less of an impact on the CA relationship since witnesses may be less inclined to choose unless they are quite confident that they can identify the suspect. Since prior research has shown that high confidence was still predictive of accuracy even after 9 months (Wixted et al., 2015), we expected high confidence judgments to be diagnostic of identification accuracy despite the negative effects of repeated exposure and retention interval. We expected the initial confidence judgment (Lineup 1) with the shortest retention interval to be associated with higher accuracy than confidence judgments made after repeated exposures (Lineup 2).

Condition	Event	Lineup 1	Lineup 2
10mins-3days	Crime Video	10 minutes	3 days
3days-3days	Crime Video	3 days	3 days

Figure 3. Experiment 1 design. The delay condition determined the length of delay between the Event and Lineup 1, but Lineup 1 and Lineup 2 were always spaced three days apart. All the participants saw the same lineup two times (Lineup 1 and Lineup 2).

In addition to the CA relationship, we explored the relationship between response latency and accuracy. Although prior studies found inconsistent optimal time-boundaries that differentiate between correct and incorrect choosers (Weber & Brewer, 2006; Weber et al., 2004), correct choosers were generally faster than incorrect choosers. However, these studies did not examine the effect of repeated identifications on response times. Therefore, like confidence judgments, the effectiveness of response time as an indicator of accuracy should also be examined in context of repeated identifications.

Moreover, we also examined consistent versus inconsistent identification decisions across repeated identifications. An example of consistent identification decisions is identifying someone in the initial identification and identifying the same person again in the subsequent identification (a consistency effect); inconsistent decisions can involve changes in identification decision across repeated lineups. For example, inconsistent witnesses might reject the initial lineup and then identify someone in the subsequent lineup (i.e. a no-to-suspect shift due to misplaced familiarity). Godfrey and Clark (2010) found that consistent witnesses who identified the same person in both the show-up and the subsequent lineup increased their confidence in their lineup identifications. In a repeated lineup study, consistent witnesses were equally confident in their initial and subsequent lineup identifications (Steblay et al., 2013); overall, inconsistent witnesses were less confident than consistent witnesses (Godfrey & Clark, 2010; Steblay et al., 2013). Based on these findings, we expected consistent witnesses to maintain or increase their confidence in the subsequent identifications and inconsistent witnesses to be less confident than consistent witnesses.

2.1 Method

2.1.1 Participants

787 participants were recruited via Amazon Mechanical Turk (MTurk), but only 591 participants completed all the sessions. Tables 1 and 2 show the number of participants in each delay condition for target-present and target-absent lineups, respectively. The remaining 196 of 787 participants consisted of those who had technical difficulties, completed the follow-up session later than the deadline, or dropped out of the study. The data from these 196 participants were not included in the following analyses. Participants were compensated with cash payment ($2.50).

2.1.2 Design

The experiment was a 2 (lineup type: target-present, target-absent) x 2 (initial delay: 10mins-3days or 3days-3days) x 2 (repeated lineup: Lineup 1, Lineup 2) mixed design. The lineup type and the initial delay condition were manipulated between-subjects, and the number of repeated lineup was manipulated within-subjects (see Figure 3).

The procedure consisted of the Event (i.e. the crime video), Lineup 1, and Lineup 2. The delay between the Event and Lineup 1 depended on the assigned initial delay condition. For example, participants in the 10mins-3days delay condition saw Lineup 1 ten minutes after watching the crime video, whereas those in the 3days-3days delay condition saw Lineup 1 three days after watching the crime video. Regardless of the delay between the Event and Lineup 1 session, the subsequent session (Lineup 2) was always three days after Lineup 1. That is, all participants saw Lineup 2 three days after Lineup 1.

In addition to the eyewitness task, participants also completed two distractors tasks (a free recall task and a lexical decision task) right after seeing the crime video. The free recall task consisted of presenting 30 words from two categorized word sets. Each word was presented for 2 seconds. Participants were given four minutes to recall as many words as possible, in any order. After this task, participants completed a lexical decision task. They were asked to determine whether a string of text was a word or nonword. They completed 16 practice trials before the 80 trials in the experimental block. Both the free recall and the lexical decision task were presented immediately following the crime video. For participants in the 10mins-3days delay condition, these two tasks were completed during the 10mins-3days delay before they saw Lineup 1. Those in the 3days-3days delay condition completed these two tasks following the crime video, but they did not see Lineup 1 until three days later.

All sessions took place online. Email reminders were sent for each subsequent session along with a link to the online experiment and participant login information. Once they received the email, they were given 36 hours to complete the follow-up session.

2.1.3 Materials

The present study used the video and lineup materials from Mickes, Flowe and Wixted (2012). The video showed a white male (the perpetrator) walking into an unoccupied office and stealing a laptop from the office desk. Viewers had a clear view of the perpetrator as he leaves the office with the laptop. The video lasted for 23 seconds and the thief's face was in view for roughly 4-5 seconds.

Participants received one of two lineups: a target-present or a target-absent lineup. Each lineup consisted of six members. The only difference between the two lineups was that the target in the target-present lineup was replaced by a filler in the target-absent lineup. The lineup photos from Mickes et al. (2012) were digitally modified so that all the lineup members had a black shirt. This step was taken to prevent participants from using the color of clothing as a cue.

The free recall task consisted of 30 words from two categorized word lists taken from the norms of Van Overschelde, Rawson and Dunlosky (2004). 15 words from the *fruit* word list and 15 words from the *vegetable* word list were used in the present study (Van Overschelde et al., 2004). The lexical decision task consisted of 16 practice trials (4 high- and 4 low-frequency words, and 8 nonwords) and 80 experimental trials (20 high- and 20 low-frequency words, and 40 nonwords). Both the words and nonwords were taken from Yap, Balota, Tse, and Besner (2008). None of the words from the free recall task overlapped with those in the lexical decision task.

2.1.4 Procedures

First, participants saw a video of the target. They were simply instructed to pay attention to the video. Immediately after the video, participants were asked to complete two distractor tasks (a free recall task and a lexical decision task). In the free recall task, they saw a total of 30 words. The order of the words was randomized. Each word was presented for two seconds followed by a blank screen for 500 ms. After the last word, participants were asked to recall as many words as possible, in any order. They were given four minutes to recall and then the screen automatically advanced to the lexical decision task. In the lexical decision task, participants were asked to determine whether each letter string was a word or nonword. There were 16 practice trials followed by the 80 experimental trials. Each trial consisted of the following sequence of events: 1) a fixation point (+) presented at the center of the screen for 400 ms. 2) a blank screen was shown for 200 ms. 3) the stimulus appeared at the location of the fixation point and remained there until the participant made a response. They were instructed to press the “M” key for words and the “Z” key for nonwords. If they made a correct response, a blank screen was presented for 1,600 ms before the fixation point appeared again. For incorrect responses, the word “Incorrect” was shown slightly below the fixation point for 1,600 ms. The free recall and the lexical decision task took at least ten minutes to complete, but the exact duration depended on the participant’s pace during the lexical decision task.

After the lexical decision task, participants saw Lineup 1 (the 10mins-3days delay condition) or were asked to wait for a follow-up email in the next few days (the 3days-3days delay condition). For Lineup 1, participants were presented with either a target-present or target-absent lineup. The exact lineup was presented in Lineup 2 (e.g. participant 1 saw the same target-present lineup for both sessions). However, the position of the lineup members was randomized between Lineup 1 and Lineup 2. Prior to seeing the lineup, participants were told that the suspect may or may not be in the lineup. They made their identification decision by clicking on the face of target, and they clicked the “Next” button to submit their choice. If participants judged that the target was not in the lineup, they could reject the lineup via the “Not Present” button. Response time tracking started when participants saw the lineup and ended when they clicked the “Next” button after selecting one of the faces or the “Not Present” button. After making their identification decision, participants were asked to make a confidence judgment using a 0 to 100 scale (0 = Not confident at all; 100 = Very confident). Participants made their confidence by moving a slider scale. The initial starting position of the confidence scale was 0. As they moved the slider, the corresponding numerical value was displayed right above the slider. Both Lineup 1 and Lineup 2 used the same procedure. No feedback was provided about whether participants made the correct choice. At the start of their follow-up session for Lineup 2, participants were again asked to identify the suspect from the video and provide a confidence judgment. After completing Lineup 2, participants were thanked for their participation and debriefed.

2.2 Results

2.2.1 Frequency of Response Types

Tables 1 and 2 show the frequency of each type of identification responses for target-present and target-absent lineups, respectively. There are three possible types of responses in target-present lineup: suspect identification (the correct response), lure identification, and not present. On the other hand, there are only two types of responses in target-absent lineups: false identification and not present (the correct response). These identification responses were also categorized as either choosers or nonchoosers. C*hoosers* for the target-present lineups consist of participants who made a suspect (i.e. a correct identification) or lure identification, whereas c*hoosers* for the target-absent lineups are those who selected any of the six lures (i.e. a false identification). *Nonchoosers* are participants who gave a “Not Present” response. In target-present lineups, a “Not Present” response is considered as a “Miss” because the target was in the lineup; in target-absent lineups, a “Not Present” response is a “Correct Rejection” because the target was not in the lineup.

2.2.2 The Effects of Repeated Lineups and Delay on Choosing

First, we examined the tendency to choose someone from a lineup. The choosing rate for target-present lineups is the total number of suspect and lure selections divided by the total number of target-present lineups; the choosing rate for target-absent lineups is the same as the false identification rate because the suspect is not in the lineup. We had several planned comparisons. For example, to examine the effects of repeated lineups, we compared Lineup 1 and Lineup 2 choosing rates. Also, we were interested if the choosing rate differed across delay

Table 1

Proportions of Identification Responses for Target-Present Lineups

		Choosers		Suspect		Lure		Not present		Total
Lineup	Delay	No.	%	No.	%	No.	%	No.	%	No.
1	10mins-3days	95	67.4	55	39.0	40	28.4	46	32.6	141
	3days-3days	91	59.1	37	24.0	54	35.1	63	40.9	154
	Overall	186	63.1	92	31.2	94	31.9	109	36.9	295
2	10mins-3days	106	75.2	63	44.7	43	30.5	35	24.8	141
	3days-3days	113	73.4	44	28.6	69	44.8	41	26.6	154
	Overall	219	74.2	107	36.3	112	38.0	76	25.8	295

Table 2

Proportions of Identification Responses for Target-Absent Lineups

		False Identifications		Correct Rejections		Total
Lineup	Delay	No.	%	No.	%	No.
1	10mins-3days	70	49.0	73	51.0	143
	3days-3days	93	60.8	60	39.2	153
	Overall	163	55.1	133	44.9	296
2	10mins-3days	95	66.4	48	33.6	143
	3days-3days	104	68.0	49	32.0	153
	Overall	199	67.2	97	32.8	296

conditions (e.g. whether the choosing rate differed for the 10mins-3days and the 3days-3days delay condition).

A separate analysis was conducted for each lineup type (target-present or target-absent), using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2017). Repeated lineups and delay conditions were the predictors; participants were specified as a random effect for both models. The status of choosing was the dependent variable (choosers were coded as 1 and nonchoosers were coded as 0). Because we were interested in specific comparisons, we constructed a no-intercept model and performed simultaneous tests for general linear hypotheses. These multiple comparisons were done with help of the multcomp package (Hothorn, Bretz, Westfall, & Heiberger, 2012). Adjusted p-values were reported for these multiple comparisons.

Target-present lineups. There was a main effect of repeated lineups, Beta = -1.03, $z = -2.77$, $p<.05$. The choosing rate increased when the lineup was repeated. From Lineup 1 to Lineup 2, the 10mins-3days delay condition increased from 67.4% to 75.2% and the 3days-3days delay condition increased from 59.1% to 73.4%. In contrast, choosing rate appeared to be lower when there was a longer delay. For example, the Lineup 1 choosing rate for the 3days-3days delay condition (59.1%) was lower than the Lineup 1 choosing rate for the 10mins-3days delay condition (67.4%). However, there was no main effect of delay. In short, regardless of delay, the choosing rate tended to increase when a lineup was repeated.

Target-absent lineups. Similarly, there was a main effect of repeated lineups, Beta = -1.04, $z = -3.04$, $p <.05$. The choosing rate also increased from Lineup 1 to Lineup 2 for target-absent lineups. The 10mins-3days delay condition increased from 49.0% to 66.4% and the 3days-3days delay condition increased from 60.8% to 68.0%. The increase in choosing rate was significant for the 10mins-3days delay condition, Beta = -.72, z = -2.98, $p < .05$, but not for the 3days-3days delay condition. Next, we examined the effect of delay. Although there was no main effect of delay, the Lineup 1 choosing rate in the 3days-3days condition appeared to be greater than the 10mins-3days condition. This was contrary to our prediction that choosing rate decreases as a function of delay. The 10mins-3days Lineup 1 choosing rate (49.0%) was lower than the 3days-3days Lineup 1 choosing rate (60.8%). One possible explanation is that participants were better able to reject a target-absent lineup after a 10-min delay than a 3-day delay due to their better memory of the suspect in the short delay condition. We will further consider this finding in the discussion section.

Interestingly, both the target-present and target-absent lineups for the 3days-3days delay conditions showed similar choosing rates (59.1% vs. 60.8%, respectively). Therefore, it appears that people were equally inclined to choose someone from a lineup after a 3-day delay regardless of whether the target was actually present in the lineup.

2.2.3 The Effects of Repeated Lineups and Delay on Accuracy

The statistical procedures were the same as those in the previous section with the exception that the dependent variable was the accuracy of identification decision (correct decisions were coded as 1 and incorrect decisions were coded as 0). The correct decision is a suspect identification. All the other possible responses in a target-present lineups (e.g. lure identification and “Not Present”) were treated as incorrect decisions. Again, the planned comparisons were the same as those in the previous section.

The number of suspect IDs appeared to increase with repeated lineups, but this increase was not significant. For the 10mins-3days delay condition, it increased from 39.0% to 44.7%. For the 3days-3days delay condition, it increased from 24.0 to 28.6%. None of these increases reached significance. On the other hand, there was a main effect of delay on suspect IDs, Beta = 1.41, z = 3.97, p<.001. Lineup 1 suspect ID rate was significantly higher in the 10mins-3days delay than the 3days-3days delay condition, Beta = .70, $z = 2.75$, $p < .05$. This difference in suspect ID rate remained significant in Lineup 2, Beta = .70, $z = 2.86$, $p < .05$. In short, the effect of delay on suspect IDs were as expected. Although suspect IDs appeared to increase with repeated lineups, this effect was not significant.

We further examined this finding by looking at the decision shifts in target-present lineups for both delay conditions (see Table 3). For both the 10mins-3days and the 3days-3days delay condition, a majority of the participants consistently picked the suspect across Lineup 1 and Lineup 2. There were some participants who initially made a lure ID or a "Not Present" response, but then picked the suspect at the subsequent lineup. For the 10mins-3days delay condition, the number of *Not Present to Suspect ID* responses was the same as the number of *Not Present to lure ID* responses, indicating a probable criterion shift. On the other hand, for the 3days-3days delay condition, the number of *Not Present to lure ID* responses was greater than *Not Present to Suspect ID* responses. In both delay conditions, the number of *Not Present to Suspect ID* responses was greater than *lure ID to Suspect ID*. These results suggest perhaps these participants may have set a higher threshold for choosing and therefore, they were initially less inclined to choose. We will further consider this finding in the discussion section.

2.2.4 Identification Decisions and Confidence Ratings

Next, we examined the average confidence ratings associated with each identification response. First, the average confidence ratings were calculated in terms of choosers and nonchoosers. Again, separate analyses were conducted for target-present and target-absent lineups. Second, we further divided the target-present choosers into two categories: correct choosers (those who made

Table 3

Target-Present Lineup Decisions: Lineup 1 to Lineup 2 Decision Shifts

		Lineup 2		
Delay	Lineup 1	Suspect	Lure	"Not Present"
10mins-3days	Suspect	49	1	5
	Lure	5	33	2
	"Not Present"	9	9	28
3days-3days	Suspect	28	5	4
	Lure	4	45	5
	"Not Present"	12	19	32

suspect IDs) and incorrect choosers (those who made lure IDs). The analyses were conducted with linear mixed-effects regression (LMER, using the lme4 package in R). The LMER results were reported in terms of F tests. Degrees of freedom were approximated with a Type II Satterthwaite approximation. All follow-up tests were performed using simultaneous tests for general linear hypotheses via the multcomp package in R.

Target-Present Lineups. Figure 4 shows the average confidence of choosers and nonchoosers in target-present Lineup 1 and Lineup 2. We conducted a 2 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. Results indicated that there was a main effects of delay, $F(1, 289.94) = 23.05$, $p < .001$. Participants in the 10mins-3days delay condition were more confident than those in the 3days-3days delay condition. There was a main effect of repeated lineups, $F(1, 294.95) = 16.24$, $p < .001$. Lineup 1 confidence ratings were generally higher than Lineup 2 confidence ratings. There was no main effect of choosing status, but there was a significant two-way interaction between repeated lineup and choosing status, $F(1,323.51) = 11.79$, $p < .001$, and a significant three-way interaction, $F(1,322.58) = 4.85$, $p < .05$. Follow-up tests of the repeated lineup x choosing status interaction revealed that nonchoosers were significantly more confident in Lineup 1 than Lineup 2, $z = 2.50$, $p < .05$, but the choosers did not differ across Lineup 1 and Lineup 2. For the follow-up tests of the three-way interaction, we examined the delay x repeated lineups interaction for each choosing status; however, neither reached significance. These results suggested that the average confidence tended to decrease as a function of delay; repeated lineups also led to a decrease in average confidence, but this dip in confidence was more apparent for nonchoosers.

Although choosers generally gave higher confidence ratings than nonchoosers, this difference was very small. This might be partially driven by the aggregation of high confidence

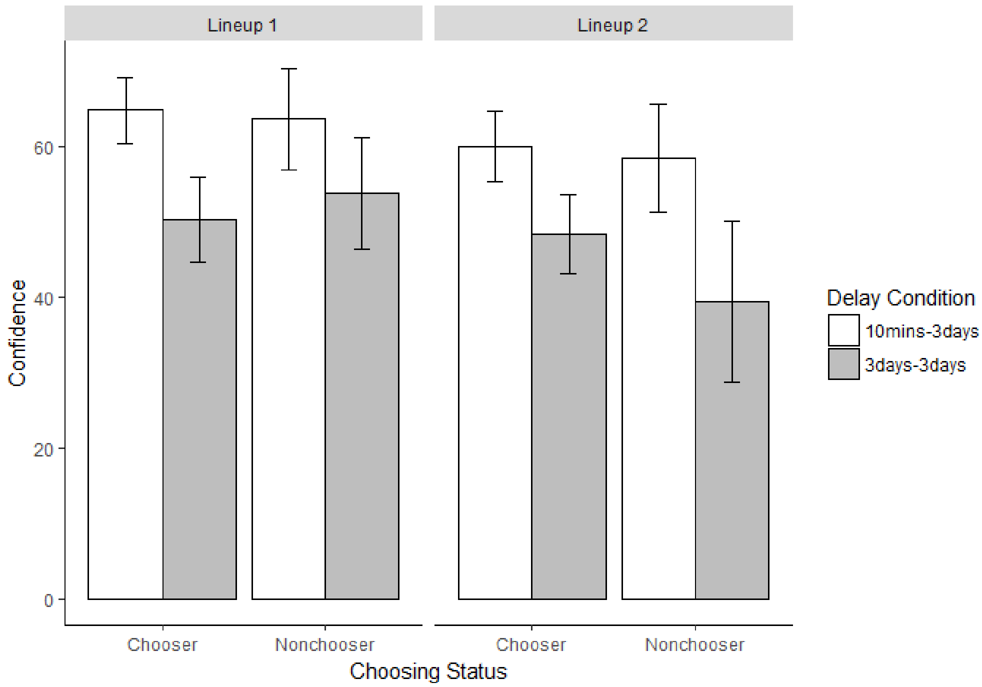


Figure 4. The average confidence of choosers vs. nonchoosers for target-present Lineup 1 and Lineup 2 in Experiment 1.

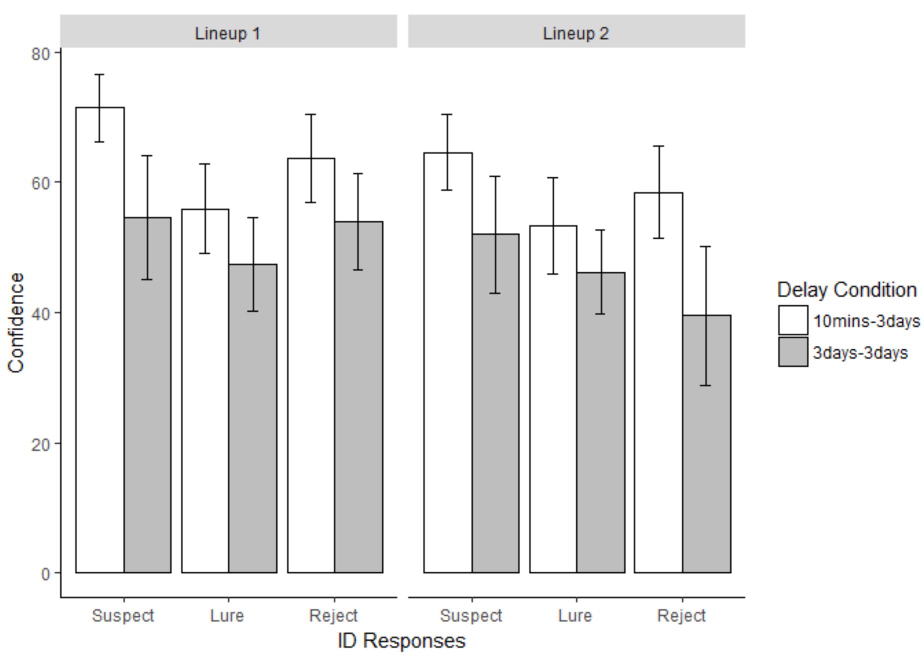


Figure 5. The average confidence of the three target-present identification responses across Lineup 1 and Lineup 2 in Experiment 1.

correct choosers (suspect identifications) and low confidence incorrect choosers (lure identifications). To examine this possibility, choosers were further divided into correct choosers (those who made suspect IDs) and incorrect choosers (those who made lure IDs). As shown in Figure 5, suspect identifications were generally made with greater confidence than incorrect identifications. This pattern was present in both Lineup 1 and Lineup 2. Therefore, the low average confidence of lure identifications was responsible for lowering the overall average confidence of choosers. Choosers were not simply more confident than nonchoosers. Correct choosers were more confident than incorrect choosers.

We conducted a 2 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 3 (identification responses) mixed model. Results indicated a main effect of delay condition, $F(1,289.96) = 20.99$, $p < .001$, and a main effect of repeated lineup, $F(1, 289.17) = 16.32$, $p < .001$. Again, participants in the 10mins-3days delay condition were more confident than those in the 3days-3days condition. Lineup 1 decisions were made with higher confidence than Lineup 2 decisions. There was no main effect of identification responses, but there was a two-way interaction between repeated lineups and identification responses, $F(2,311.47) = 5.96$, $p<.05$. Follow-up tests were conducted for the repeated lineups x identification responses interaction. In Lineup 1, the confidence ratings of suspect IDs were significantly higher than lure IDs, $z = 2.95$, $p<.05$, but the confidence ratings of suspect IDs were not significantly different from “Not Present” responses. In Lineup 2, although the confidence ratings of suspect IDs remained higher than both lure IDs and “Not Present” responses, their differences were not significant. In other words, the differences in confidence between correct and incorrect choosers became less apparent at the subsequent lineup.

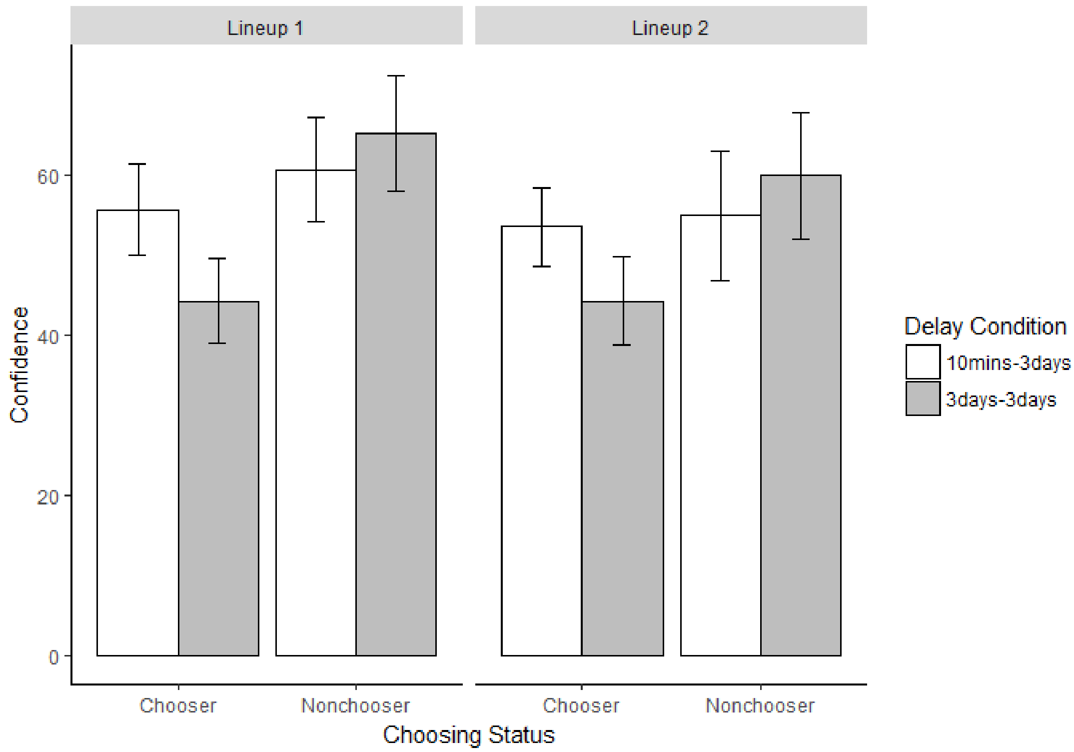


Figure 6. The average confidence of choosers vs. nonchoosers for target-absent Lineup 1 and Lineup 2 in Experiment 1.

Target-Absent Lineups. Figure 6 shows the average confidence of choosers and nonchoosers in target-absent Lineup 1 and Lineup 2. Because the target was not in target-absent lineups, all the choosers made an incorrect response. The average confidence of choosers and nonchoosers was submitted to a 2 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. Results indicated a main effect of repeated lineups, $F(1,299.94) = 3.88$, $p<.05$, and a main effect of choosing status, $F(1,562.13) = 23.62$, $p<.001$. Again, Lineup 1 decisions were generally made with higher confidence than Lineup 2 decisions. Nonchoosers (i.e. those who made a correct rejection) were more confident than choosers. There was also a two-way interaction between the delay conditions and choosing status, $F(1,561.91) = 9.34$, $p <.05$, and a two-way interaction between repeated lineups and choosing status, $F(1,337.90) = 7.24$, $p <.05$. Follow-up tests were performed for the delay x choosing status interaction. Choosers in the 10mins-3days delay condition were significantly more confident than those in the 3days-3days delay condition, z = 3.67, p <.001. However, the confidence of nonchoosers did not differ across delay conditions. Next, the follow-up tests were conducted on the repeated lineup x choosing status interaction. Nonchoosers were significantly more confident than choosers in both Lineup 1, $z = 4.15$, $p <.001$, and Lineup 2, $z = 2.59$, $p <.05$. These results suggested that the length of delay affected choosers more than nonchoosers; nonchoosers (those who made a correct rejection) were more confident than choosers, although this difference was more apparent in Lineup 1 than Lineup 2.

2.2.5 Confidence-Accuracy Relationship

The 0 to 100 confidence scale was collapsed into five confidence bins (0-20%, 21-40%, 41-60%, 61-80%, and 81-100%). Table 4 shows the distribution of suspect and false IDs across these five confidence bins (0-20%, 21-40%, 41-60%, 61-80%, and 81-100%). Because there

Table 4

The Number of Suspect and False IDs across Confidence Bins

		Lineup 1 Confidence					Lineup 2 Confidence				
Delay	Responses	0-20	21-40	41-60	61-80	81-100	0-20	21-40	41-60	61-80	81-100
10mins-3days	Suspect IDs	1	2	13	19	20	4	5	14	22	18
	False IDs	5	15	17	23	10	9	20	30	22	14
3days-3days	Suspect IDs	5	7	9	8	8	6	16	5	6	11
	False IDs	23	25	16	21	8	29	24	15	22	14

were only a few responses in the lower confidence bins, we further collapsed these five confidence bins into three confidence bins (0-40%, 41-89%, and 90-100%) to improve the stability of the calibration curves. Figure 7 shows the calibration plots for Lineup 1 and Lineup 2. For each confidence bin, the accuracy was calculated by the number of correct suspect IDs / (correct suspects IDs + false IDs) for each confidence bin. Following the calibration plot procedure from Wixted et al. (2015), the lure (or foil) identifications in target-present lineups were excluded from the present calibration plots. In other words, the number of false identifications comes from the number of incorrect identifications made in target-absent lineups. Because there was no designated innocent suspect in the target-absent lineups, the sum of false IDs for a particular confidence bins (e.g. all the false identifications made with a confidence rating of 0-40%) was divided by the lineup size (Wixted et al., 2015).

Regardless of delay conditions, both Lineup 1 and Lineup 2 revealed a positive relationship between confidence and accuracy. Accuracy reached its peak at the highest confidence bin (90-100%) in both delay conditions and lineups (Lineup 1 and Lineup 2). Overall, it appears that the confidence-accuracy relationship remained intact despite of factors such as delay and repeated lineups.

In addition to constructing calibration plots for choosers, we also created calibration plots for nonchoosers. The equation used is similar to the calculation for choosers [correct nonchoosers/(correct nonchoosers + incorrect nonchoosers)]. Figures 8 shows the nonchooser calibration plots for Lineup 1 and Lineup 2. Unlike the calibration plots for choosers, there was not a clear relationship between confidence and accuracy. Consistent with prior literature (Tekin & Roediger, 2017), the calibration curves for nonchoosers tended to be flat.

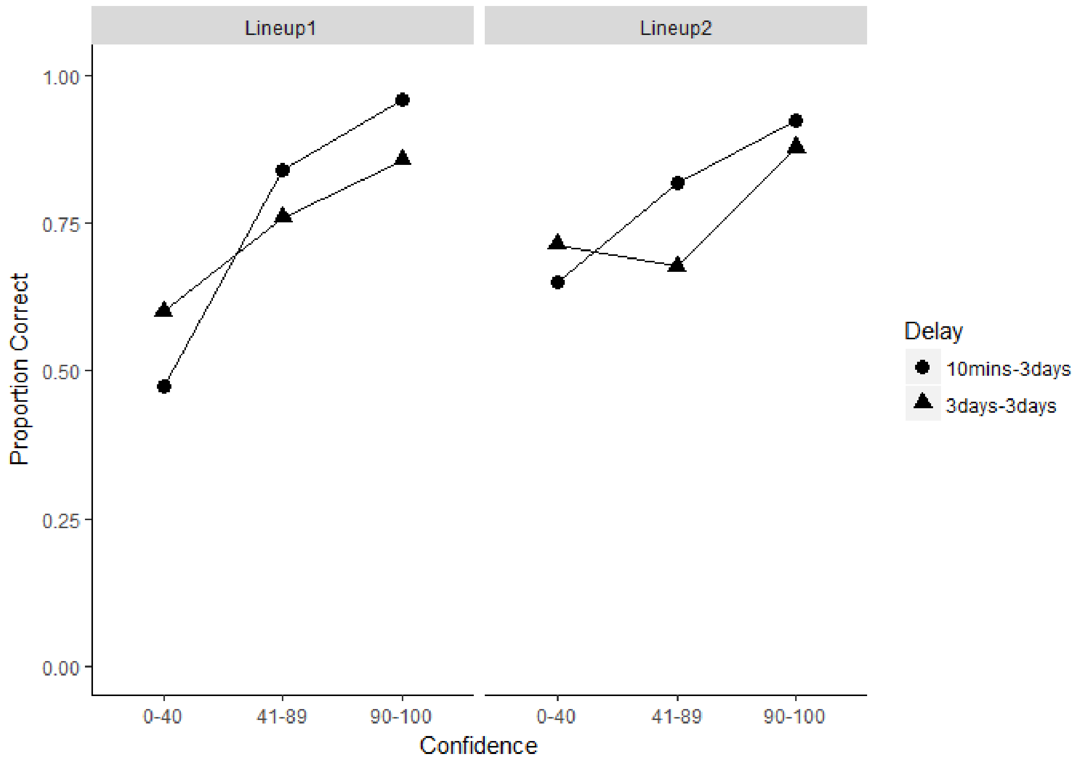


Figure 7. Experiment 1 calibration plots for choosers in Lineup 1 and Lineup 2.

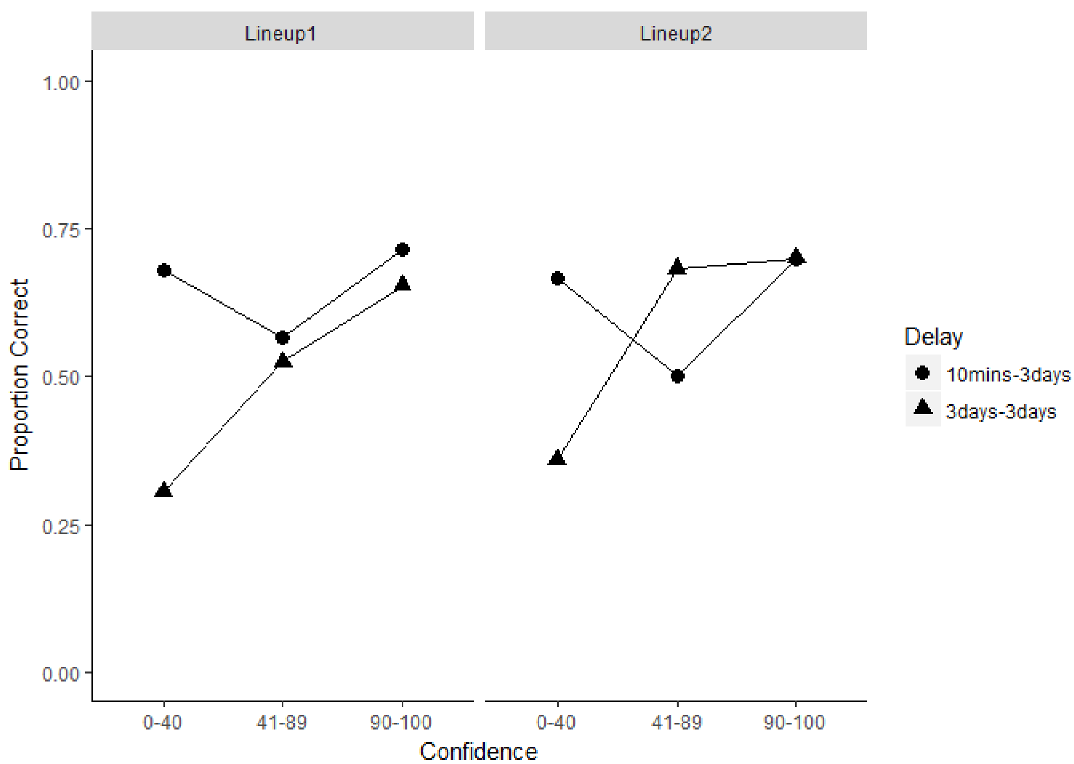


Figure 8. Experiment 1 calibration plots for nonchoosers in Lineup 1 and Lineup 2.

2.2.6 Confidence Ratings and Response Times as Indicators of Identification Accuracy

Next, we compared the effectiveness of confidence ratings and response times as indices of identification accuracy. In the previous sections, we examined the average confidence associated with accurate identification decisions. Analyses for average response times were also conducted. For readers interested in the response time data, we analyzed average response time (see Appendix A) and the optimal time-boundary (see Appendix B). Averages, however, only tell one side of the story. To effectively use these indicators of accuracy in real-life settings, we would need to have a reliable cutoff or criterion (e.g. a confidence rating of 90% or above). Thus, the primary purpose of this section was to evaluate the utility of these indicators based on a specific cutoff.

To compare the predictability of confidence ratings and response times as indicators of accuracy, we examined the number of suspect IDs and the false alarms made with high confidence (90% or above) or a fast response time (10 seconds or faster). The 10 second cutoff was based on Dunning and Perretta (2002)'s claim that the optimal time-boundary that best differentiates correct choosers from incorrect choosers is 10-12 seconds. Weber et al. also examined their results using this cutoff. In addition, Weber et al. also examined accuracy for responses made with a confidence rating of 90% or above. For these reasons, we decided to examine our data based on these combined criteria.

The calculation of accuracy was the same as the calculation for calibration [correct suspect IDs / (correct suspects IDs + incorrect suspect IDs)]. Because there was no designated innocent suspect in the target-absent lineups, the sum of incorrect suspect IDs for a confidence bin (rating of 90% or above) or a RT bin (10 seconds or earlier) was divided by the number of lineup members. Tables 5 and 6 show the accuracy for high confidence and fast response time,

Table 5

Accuracy of Responses Above or Below 90% Confidence

Lineup	Delay	Confidence	% correct
Lineup 1	10mins-3days	Above	96.0
		Below	79.4
	3days-3days	Above	85.7
		Below	69.0
Lineup 2	10mins-3days	Above	92.3
		Below	78.4
	3days-3days	Above	87.8
		Below	69.7

Table 6

Accuracy of Responses Before or After 10 seconds

Lineup	Delay	Response Time	% correct
Lineup 1	10mins-3days	Before	87.6
		After	75.9
	3days-3days	Before	58.3
		After	74.1
Lineup 2	10mins-3days	Before	81.5
		After	77.4
	3days-3days	Before	67.7
		After	75.8

Table 7

Accuracy of Responses with High Confidence and Fast Response Time

Lineup	Delay	% correct
Lineup 1	10mins-3days	97.1
	3days-3days	60.0
Lineup 2	10mins-3days	92.3
	3days-3days	85.7

respectively. Regardless of delay or repeated lineups, high confidence generally yielded high accuracy compared to response time. Based on these criteria, confidence ratings appeared to be a more dependable indicator of accuracy compared to response time. In Table 7, we calculated the accuracy based on decisions made with a combination of a high confidence rating (90% or above) and a fast response time (10 seconds or sooner). In some cases, accuracy was slightly improved using both the confidence and response time cutoff; however, it did not always increase accuracy. For instance, the accuracy of Lineup 1 for the 3days-3days condition was 85.7% with the confidence cutoff, but it was 60.0% with a cutoff based on both high confidence and fast response time. In other words, people who were highly confident may not always be the same people who were quick to respond. However, the accuracy from the combination of these two measures were higher than a RT criterion alone. In sum, these results suggest that confidence was a better indicator of accuracy compared to RT.

2.2.7 Consistent and Inconsistent Decisions across Lineup 1 and 2

Table 8 shows the proportion of same and different decisions across Lineup 1 and 2. The same decision category consists of identification decisions that were the same across Lineup 1 and Lineup 2 (e.g. selecting the suspect and selecting the suspect again or selecting a lure and selecting the same lure again); the different decision category includes identification decisions that differed across Lineup 1 and Lineup 2 (e.g. selecting a lure and selecting another lure, rejecting a line and then choosing someone from a lineup, or vice versa).

Overall, participants generally made the same decisions across the lineups. For target-present lineups, 70.9% of the participants in the 10mins-3days delay condition and 64.3% of the participants in the 3days-3days delay condition maintained the same identification decision across repeated lineups. For target-absent lineups, 60.8% of the participants in the 10mins-3days

Table 8

Proportion of Same and Different Decisions across Lineups 1 and 2

Lineup Type		Same Decision	Different Decision
Target-Present	10mins-3days	70.9	29.1
	3days-3days	64.3	35.7
Target-Absent	10mins-3days	60.8	39.2
	3days-3days	64.1	35.9

Table 9

Confidence Ratings of Same and Different Decisions across Lineups 1 and 2

		Same		Different	
Lineup Type	Delay	Lineup 1	Lineup 2	Lineup 1	Lineup2
Target-Present	10mins-3days	67.0	63.7	58.4	49.9
	3days-3days	56.8	51.8	42.8	35.8
Target-Absent	10mins-3days	56.8	56.4	60.4	50.3
	3days-3days	52.2	51.1	52.9	46.1

delay condition and 64.1% of the participants in the 3days-3days delay condition kept the same identification decision across repeated lineups.

Next, we also examined the average Lineup 1 and 2 confidence ratings associated with these same and different decisions (see Table 9). Regardless of whether participants maintained the same identification decisions or made different decisions across the lineups, they tended to lower their confidence ratings at Lineup 2. This pattern was observed in both target-present and target-absent lineups. In short, participants typically repeated their Lineup 1 decision in Lineup 2; however, they decreased their Lineup 2 confidence ratings regardless of whether they were making the same decision or a different decision at Lineup 2.

2.2.8 Lineup 1 and 2 Decision Combinations and Their Confidence Ratings

Next, we was further divided these same and different decisions into subgroups of interest (see Table 10). For target-present lineups, we were interested in three possible Lineup 1 and 2 decision combinations: repeated suspect IDs, repeated lure IDs, and repeated “Not Present” responses. For target-absent lineups, we were interested in three possible Lineup 1 and 2 decision combinations: repeated lure IDs, repeated “Not Present” responses, and a “Not Present” response to choosing someone. We examined the proportion of these decision combinations and their average confidence ratings.

Target-Present Lineups. Out of the three types of decision combinations, suspect IDs were the most repeated decision combination. For the 10mins-3days delay condition, 55 suspect IDs were made in Lineup1 and 49 of these suspect IDs (89.1%) were repeated in Lineup 2. For the 3days-3days delay condition, 37 suspect IDs were made in Lineup 1 and 28 of these suspects IDs (75.7%) were repeated in Lineup 2. However, there was also a sizable proportion in the other two types of decision combinations. For instance, 23 of the 40 lure IDs (57.5%) in the 10mins-

Table 10

The Count and Confidence Ratings of Decision Types across Lineups 1 and 2

			No. of Decisions			Confidence Ratings	
Lineup Type		Response Type	Lineup 1	Lineup 2	%	Lineup 1	Lineup 2
Target-Present	10mins-3days	Repeated Suspect IDs	55	49	89.1	72.3	69.3
		Repeated Lure IDs	40	23	57.5	61	59.9
		Repeated "Not Present"	46	28	60.9	62.8	57.1
		"Not Present" to Choosing	46	18	39.1	65.1	51.4
	3days-3days	Repeated Suspect ID	37	28	75.7	60.5	60.2
		Repeated Lure ID	54	39	72.2	53	54.5
		Repeated "Not Present"	63	32	50.8	58.3	41.2
		"Not Present" to Choosing	63	31	49.2	49.5	36.6
Target-Absent	10mins-3days	Repeated Lure ID	70	41	58.6	54.6	58.7
		Repeated "Not Present"	73	40	54.8	59.4	54.4
		"Not Present" to Choosing	73	33	45.2	62.1	50.7
	3days-3days	Repeated Lure ID	93	56	60.2	42.3	46.2
		Repeated "Not Present"	60	37	61.7	68.8	59
		"Not Present" to Choosing	60	23	38.3	59.3	40.6

Note. The percentages listed under “No. of Decisions” were the proportion of each response type.

3days delay condition and 39 of the 54 lure IDs (72.2%) in the 3days-3days delay condition were kept in the second lineup. 28 out of 46 (60.9%) “Not Present” responses were repeated in the 10mins-3days delay condition; 32 out of 63 (50.8%) “Not Present” responses were maintained in the 3days-3days delay condition. In short, a portion of participants continued to select the wrong person or incorrectly reject the lineup. Regardless of the types of decision combinations, there was a general tendency to lower confidence in Lineup 2.

Target-Absent Lineups. Similarly, the Lineup 1 decisions made in target-absent lineups were also often repeated in Lineup 2. 41 of 70 (58.6%) lure IDs in the 10mins-3days delay condition and 56 of 93 (60.2%) lure IDs in the 3days-3days delay condition were maintained across repeated lineups. There was a considerable number of participants who were able to correctly reject the lineup across repeated lineups. For instance, 40 of the 73 (54.8%) “Not Present” responses in the 10mins-3days delay condition and 37 of 60 (61.7%) “Not Present” responses in the 3days-3days delay condition were kept. However, there was also a sizeable proportion of “Not Present” response in Lineup 1 to choosing in Lineup 2 decisions in both the 10mins-3days delay (45.2%) and the 3days-3days delay (38.3%) conditions. This shift may be due to misplaced familiarity in Lineup 2, because all faces had been seen three days previously. Therefore, both correct decisions (rejecting a target-absent lineup) and incorrect decisions (selecting the same lure) were likely to be repeated, although there was a tendency to shift from a correct rejection toward an incorrect choice when a lineup was repeated. Similar to the confidence ratings in target-present lineups, confidence ratings decreased with repeated lineups across the different decision combinations in target-absent lineups, with the exception of repeated lure IDs. For both delay conditions, there was a minor increase in confidence for repeated lure IDs.

Overall, participants often repeated their decisions regardless of whether their initial decision was correct. Contrary to our predictions, participants did not increase their confidence when they repeated the same decisions across Lineups 1 and 2, with the exception of aforementioned repeated lure IDs in target-absent lineups.

2.2.9 Confidence and Decision Consistency

Although participants generally made the same decisions in both Lineup 1 and Lineup 2, could we predict the likelihood of repeating a decision using the Lineup 1 confidence rating? In other words, if people were confident in their Lineup 1 decision (correct or incorrect), would these same people repeat their Lineup 1 decision in Lineup 2 more often than those who were less confident in their Lineup 1 decisions? To investigate this predictability of Lineup 1 confidence, we categorized participants by whether they made a consistent or inconsistent decision across Lineup 1 and Lineup 2. We examined the proportion above and below each confidence level (20%, 40%, 60%, 80%, and 100%) to better gauge the effects of confidence and reproducibility of decisions across lineups.

Confidence Levels. Using the level of confidence participants expressed in Lineup 1, we calculated the proportion of these participants who repeated their decisions in Lineup 2. Table 11 shows the proportion of consistent responses above and below each confidence level for each delay condition. The "Confidence" column indicates the level of confidence expressed at Lineup 1. For instance, when the confidence level is at 40, the above column shows the proportion of consistent responses with a confidence rating of 40 to 100, whereas the below column shows the proportion with a confidence ratings of 0 to 39. Because nothing is above a confidence rating of 100, the above column shows the proportion of consistent responses with a confidence rating of 100 and the below columns shows the proportion with a confidence rating of 0 to 99. Similarly,

Table 11

The Proportion of Consistent Responses Above and Below each Confidence Level

Delay	Confidence	Above	Below
10mins-3days	0	65.9	NA
	20	65.8	66.7
	40	65.2	68.5
	60	69.4	60.5
	80	**69.7**	64.4
	100	**90.9**	64.8
Delay	Confidence	Above	Below
3days-3days	0	64.2	NA
	20	64.2	64
	40	66.2	60.7
	60	70.3	59.2
	80	**76.5**	60.7
	100	**87.5**	62.9

Note. The "confidence" column indicates the level of confidence expressed at Lineup 1. Using the level of confidence participants expressed in Lineup 1, we calculated the proportion of these participants who repeated their decisions in Lineup 2.

because nothing is below a confidence rating of 0, the above column simply shows the proportion of consistent responses with a confidence rating of 0 to 100.

Regardless of delay conditions, high confident Lineup 1 responses were more likely to be repeated. For the 10mins-3days delay condition, 69.7% of the consistent decisions were made with a Lineup 1 confidence rating of 80 or higher. In the 3days-3days delay condition, 76.5% of the consistent decisions had a Lineup 1 confidence rating of 80 or higher. Although it was not a perfect relationship, Lineup 1 confidence ratings were generally predictive of whether a decision would be repeated.

2.3 Discussion

Experiment 1 attempted to disentangle the effects of repeated identification procedures and delay on eyewitness identification. Consistent with my first-year project results, the present experiment showed that repeated lineups increased people's tendency to choose. Aside from repeated lineups, the length of delay also exerted effects on choosing behaviors. For target-present lineups, choosing rate decreased as the length of delay increased, although the decrease was not significant. However, the opposite pattern was observed for target-absent lineups. Notably, when Lineup 1 occurred 3 days after watching the crime video, participants were equally likely to choose someone from a target-present lineup (59.1%) and a target-absent lineup (60.8%). On the other hand, when Lineup 1 was presented 10 minutes later, the choosing rate for target-absent lineups was at 49.0%, whereas the choosing rate for target-present lineups was 67.4%. The lower target-absent choosing rate in the short delay condition may reflect a more intact memory of the suspect from the video, which it in turn allowed people to correctly reject a target-absent lineup.

Although the number of suspect IDs declined as a function of delay, repeated lineups actually led to an apparent but non-significant increase in suspect IDs. This finding, however, is not unique to the present experiment as Steblay et al. (2013) reported similar findings. In their experiment, they also found that some participants who did not initially make a suspect ID did so in the subsequent lineup. Steblay et al. suggested that this decision pattern was due to misplaced familiarity because there were also participants who only made an innocent suspect ID at Lineup 2. Consistent with this explanation, the present study also showed that participants who initially did not make a false ID did so in Lineup 2. Another explanation Steblay et al. provided was that participants may not be certain in the initial lineup, but applied a "close enough" judgment in the subsequent lineup. Because we did not ask participants to explain their identification decisions, we do not know why these participants responded the way they did. Furthermore, the increase in suspect IDs was not significant. Nonetheless, we will see whether this effect replicates in Experiment 2.

Another question of interest is the effects of repeated lineups and delay on identification confidence. We will first discuss the average confidence results for target-present lineups followed by target-absent lineups. For target-present lineups, participants in the long delay condition were generally less confident in their identification decisions compared to those in the short delay condition. Regardless of delay or decision types, participants were generally less confident in Lineup 2 compared to Lineup 1. For the target-present lineups, choosers were not significantly more confident than nonchoosers. This small difference can be explained by the aggregation of high confidence correct choosers and low confidence incorrect choosers. Consistent with prior literature and our prediction, correct choosers (i.e. people who made suspect IDs) tended to more confident than incorrect choosers (i.e. people who selected a lure).

While suspect IDs were made with significantly higher confidence than lure IDs in Lineup 1, this difference was not significant in Lineup 2. That is, the difference in confidence between correct choosers and incorrect choosers became less distinguishable in Lineup 2.

For target-absent lineups, nonchoosers (those who made a correct rejection) were generally more confident than choosers (those who made a false ID). Choosers in the 10mins-3days delay condition were more confident than those in the 3days-3days delay condition; nonchoosers did not differ across delay condition. In other words, even though choosers were clearly making a mistake, their confidence level was linked to the length of delay. Furthermore, nonchoosers were generally more confident than choosers in both Lineup 1 and Lineup 2, but the difference in average confidence was greater in Lineup 1 than Lineup 2.

In short, correct decisions (suspect IDs or correct rejections) were also typically made with higher confidence than incorrect decisions (lure IDs in target-present lineups or false IDs in target-absent lineups). However, the effects of delay and repeated lineups should be taken into consideration because both factors influence confidence level. For instance, incorrect choosers were more confident in their decisions when the delay was short compared to long. Therefore, the incorrect choosers' high confidence here reflected more an effect of delay rather than accuracy. Furthermore, repeated lineups seemed to reduce the difference in confidence between correct and incorrect decisions. In other words, it is more difficult to distinguish correct decisions from incorrect ones in a subsequent lineup on the basis of confidence.

Despite of the negative effects of repeated lineups and delay on eyewitness identification, the confidence-accuracy relationship for choosers remained intact. As predicted, the confidence-accuracy relationship was better calibrated for Lineup 1 than Lineup 2. The 10mins-3days delay condition tended to have higher accuracy than the 3days-3days delay condition. However, regardless of repeated lineups or delay effects, high confidence was associated with high accuracy. On the other hand, the calibration curves of nonchoosers did not provide a meaningful relationship between confidence and accuracy. Overall, these results suggest that the confidence-accuracy relationship for choosers is relatively robust.

Furthermore, when we examined confidence judgments and response times based on specific criteria, we found that the accuracy was higher for high confidence responses (90% or above) compared to fast response time (10 seconds or earlier). The combination of high confidence and fast response time was also superior to a fast RT criterion; however, it was not always better than high confidence criterion. Furthermore, consistent with Brewer and Wells (2004), our time-boundary analysis also revealed that the optimal time-boundary varied across different manipulations (see Appendix B). That is, although fast RTs are associated with correct decisions, it is difficult to establish a specific criterion that best differentiates the correct responses from the incorrect ones. Therefore, compared to response times, confidence judgements appeared to be a better and more stable indicator of identification accuracy.

Next, we will discuss the consistency of identification decisions across repeated lineups. In more than 50% of the cases, participants repeated their Lineup 1 decisions in Lineup 2. This pattern was observed across delay conditions and lineup types. Regardless of whether people repeat the same response or make a different response, they were generally less confident in their Lineup 2 decisions than for Lineup 1. We further divided these repeated identification decisions by categories of interest such as repeated suspect IDs, lure IDs, or “Not Present” responses. Contrary to our predictions, repeated lineups did not lead to confidence inflation. In fact, it appears that when people made the same decisions across lineups (e.g. choose the suspect at Lineup 1 and 2), their confidence at Lineup 2 were, on average, lower than Lineup 1. There were a few exceptions. When people selected a lure from a target-absent lineup and then selected the same lure at the subsequent lineup, they did increase their confidence at Lineup 2. The increase was minimal (less than 5%); therefore, this increase may simply be noise in the data. We will examine whether this pattern is replicated in Experiment 2.

Although majority of the participants repeated their decisions, we found that the likelihood of repeating a decision could be predicted by the Lineup 1 confidence ratings. We examined Lineup 1 confidence ratings by confidence level (20%, 40%, 60%, 80%, and 100%). When people gave a higher confidence rating at Lineup 1, they were more likely to repeat their decision in Lineup 2, regardless whether their Lineup 1 decision was correct or incorrect. This finding is consistent with Korat (2012)'s idea that confidence is a predictor or reproducibility.

In sum, both the effects of repeated lineups and delay influence identification performance. People performed worse when the delay was long; they were also likely to repeat the same decision in a subsequent lineup. Despite the negative effects of repeated lineups and delay, the confidence-accuracy relation remained intact. High confidence was associated with high accuracy. Confidence ratings were also predictive of whether people repeat the same decision in a subsequent lineup. Although response time is a potential post-dictive indicator of accuracy, it is less reliable than confidence. The accuracy of a high confidence criterion or a combination of high confidence and fast RTs was superior to a fast RT criterion alone.

Chapter 3: Experiment 2

The primary goal of Experiment 2 was to examine whether the length of delay between the initial identification and the subsequent identification influence how eyewitnesses make their subsequent identification decisions. For example, Godfrey and Clark (2010) found a mix of consistent identification decisions (e.g. selecting the same person in both identifications) and no-to-suspect shift (i.e. rejecting the initial identification but choosing someone in the subsequent identification) when the time interval between the show-up and subsequent lineup was one week, but they found some consistent identification decisions and few no-to-shifts when the time interval was 30 minutes. These decision pattern differences could be due to the length of delay between the two identifications. Steblay et al. (2013) used a 2-week delay between two consecutive lineups and found both consistent decisions and no-to-suspect shifts. It is uncertain, however, what decision patterns would be observed if the delay between the two consecutive lineups was longer (or shorter) than two weeks. Furthermore, both Godfrey and Clark (2010) and Steblay et al. (2013) presented their initial identification immediately after the study phase. In short, these studies neither manipulated the *initial delay* between the study phase (e.g. the "event" and initial identification nor the *subsequent delay* between two identifications. Although Pezdek and Blandon-Gitlin (2005) did examine the effects of this subsequent delay, their Lineup 1 was always a target-absent lineup. In contrast, we wanted to examine this effect when the same lineup (a target-present or a target-absent lineup) was presented twice. Therefore, more extensive studies are necessary to capture real-world forensic situations where the length between these two types of delays vary.

Experiment 2 consisted of three parts: Event, Lineup 1, and Lineup 2. The delay between the Event and Lineup 1 was the *initial delay* and the delay between Lineup 1 and Lineup 2 was the *subsequent delay*. The goals of Experiment 2 were to address the following questions: 1) Does the subsequent delay affect the identification decision made at Lineup 2? 2) Are witnesses more likely to make the same identification decision at both Lineup 1 and Lineup 2 when the subsequent delay is short (a consistency effect)? 3) Are witnesses more likely to make more inconsistent decisions (e.g. no-to-suspect shifts) when the subsequent delay is long (possibly due to misplaced familiarity)? 4) If the length of the subsequent delay affects the identification decision at lineup 2, is this effect also conditional on the initial delay length? For example, would witnesses in a long-initial and long-subsequent delay condition change their lineup decisions more often compared to those in the short-initial and short-subsequent delay, short-initial and long-subsequent, or long-initial and short-subsequent delay conditions? 5) Lastly, how does the subsequent delay affect the effectiveness of confidence judgment and response time as post-dictive indicators of accuracy?

Our general predictions for Experiment 2 were similar to those in Experiment 1; however, we also had specific hypotheses for the aforementioned research questions. We hypothesized that there would be more consistent identification decisions when the subsequent delay was short, and more inconsistent identification decisions when subsequent delay was long. For example, when the subsequent delay is short, participants might be more likely to remember the face they had selected during Lineup 1, which might compel them to pick the same face again (a consistency effect). When the subsequent delay is long, however, participants might forget their identification decision at Lineup 1 and therefore make a different identification decision at Lineup 2 (e.g., pick a filler in Lineup 1 and pick a different filler in Lineup 2). Furthermore, when the subsequent delay is long, nonchoosers (i.e. participants who rejected the lineup) in Lineup 1 might become choosers in Lineup 2 due to misplaced familiarity. Therefore, we expected to see more inconsistent identification decisions when the subsequent delay is long. We hypothesized that the long-initial and long-subsequent delay condition (the 5days-5days condition) would show the most inconsistent identification decisions compared to the other three conditions because both the long-initial and long-subsequent delay would allow for a considerable amount of forgetting to occur.

Once again, we hypothesized the confidence judgment made at Lineup 1 was more likely to be a better indicator of accuracy than the confidence judgment made at Lineup 2 because errors made at Lineup 1 were less likely to be corrected at Lineup 2 due to the consistency effect. We hypothesized the confidence-accuracy relationship to remain robust despite the varying initial and subsequent delay. Again, we expected confidence to be more reliable than response time as indicators of accuracy.

3.1 Method

3.1.1 Participants

1158 Participants were recruited via Amazon Mechanical Turk (MTurk). Only 883 of them completed all the follow-up sessions. Tables 12 and 13 show the number of participants in each delay condition for target-present and target-absent lineups, respectively. The remaining 275 participants consisted of those who had technical difficulties, completed the follow-up session later than the deadline, or dropped out of the study. The data from these 275 participants were not included in the following analyses. Participants were compensated with cash payment ($2.50).

3.1.2 Design

Similar to Experiment 1, Experiment 2 consisted of the Event, Lineup 1, and Lineup 2. Unlike Experiment 1, Experiment 2 systematically manipulated both the initial delay (i.e. the

Table 12

Proportions of Identification Responses for Target-Present Lineups

		Choosers		Suspect		Lure		Not present		Total
Lineup	Delay	No.	%	No.	%	No.	%	No.	%	No.
1	10mins-1day	84	81.6	41	39.8	43	41.7	19	18.4	103
	10mins-5days	87	81.3	52	48.6	35	32.7	20	18.7	107
	5days-1day	80	72.1	29	26.1	51	45.9	31	27.9	111
	5days-5days	79	65.3	23	19.0	56	46.3	42	34.7	121
	Overall	330	74.7	145	32.8	185	41.9	112	25.3	442
2	10mins-1day	91	88.3	40	38.8	51	49.5	12	11.7	103
	10mins-5days	93	86.9	50	46.7	43	40.2	14	13.1	107
	5days-1day	90	81.1	24	21.6	66	59.5	21	18.9	111
	5days-5days	86	71.1	24	19.8	62	51.2	35	28.9	121
	Overall	360	81.4	138	31.2	222	50.2	82	18.6	442

Table 13

Proportions of Identification Responses for Target-Absent Lineups

		False Identifications		Correct Rejections		Total
Lineup	Delay	No.	%	No.	%	No.
1	10mins-1day	70	64.2	39	35.8	109
	10mins-5days	62	59.6	42	40.4	104
	5days-1day	59	54.1	50	45.9	109
	5days-5days	74	62.2	45	37.8	119
	Overall	265	60.1	176	39.9	441
2	10mins-1day	84	77.1	25	22.9	109
	10mins-5days	75	72.1	29	27.9	104
	5days-1day	79	72.5	30	27.5	109
	5days-5days	100	84.0	19	16.0	119
	Overall	338	76.6	103	23.4	441

delay between the Event and Lineup 1) and the subsequent delay (i.e. the delay between the Lineup 1 and Lineup 2). The experiment was a 2 (lineup type: target-present, target-absent) x 4 (initial-subsequent delay: 10 minutes and 1 day, 10 minutes and 5 days, 5 days and 1 day, 5 days and 5 days) x 2 (number of lineups: Lineup 1, Lineup 2) mixed design (see Figure 9). The lineup type and the initial-subsequent delay conditions were manipulated between-subjects while the number of repeated lineups was manipulated within-subjects.

In sum, there were four variations of delay conditions: 10mins-1day, 10mins-5days, 5days-1day and 5days-5days. Participants were randomly assigned to one of these four delay conditions. In addition, they either received a target-present or target-absent lineup. There were eight conditions (four delay conditions for the target-present lineup and four delay conditions for the target-absent lineup). All the sessions took place online. Upon the completion of the first session, participants were presented with the login info, and the date for the follow-up session. In addition, email reminders were also sent for each subsequent session. Participants were given 48 hours to complete each follow-up sessions.

3.1.3 Materials

The materials for Experiment 2 were the same as those used in Experiment 1 except another filler was randomly selected to replace the target in the target-absent lineup. This was to see whether using another target-replacement filler would produce different frequencies in filler selections.

3.1.4 Procedures

The procedure was the same as Experiment 1 with the exception of varying length of initial and subsequent delay, a different target-replacement filler, and a “first-choice” rule instead of a “final-choice” rule. All participants first saw a video of the target followed by two distractor

Condition	Event	Initial Delay	Lineup 1	Subsequent Delay	Lineup 2
10mins-1day	Crime Video	10 Minutes	Lineup 1	1 Day	Lineup 2
10mins-5days	Crime Video	10 Minutes	Lineup 1	5 Days	Lineup 2
5days-1day	Crime Video	5 Days	Lineup 1	1 Day	Lineup 2
5days-5days	Crime Video	5 Days	Lineup 1	5 Days	Lineup 2

Figure 9. Experiment 2 design. Participants always saw the same lineup two times, but the length of initial delay and the length of subsequent delay varied depending on the assigned condition.

tasks (a free recall task and then a lexical decision task). Depending on their assigned condition, they either saw the initial lineup after completing the lexical decision task or 5 days later. Lineup 2 was either one day or five days after Lineup 1. For example, some participants saw Lineup 1 10 minutes after the video and they received Lineup 2 a day later. Other participants saw Lineup 1 10 minutes after the video, but they received Lineup 2 five days later. Similarly, those in the 5-day initial delay condition could see Lineup 2 a day later or five days later. Participants were given the login info (experiment access code) and the date to complete the follow-up sessions. They had a 48-hour window to complete the follow-up on their assigned date. The exact 48-hour window (e.g. the date and time) was provided for each participant upon their completion of session 1. They also received 2 email reminders for each subsequent session.

For Lineup 1 and Lineup 2, participants saw a lineup and were asked to identify the suspect. Again, they were told that the suspect may or may not be in the lineup. However, we changed the way participants submitted their identification responses. Unlike Experiment 1, choosers in Experiment 2 did not have to click on a lineup member's photo and the "Next" button to submit their response. That is, Experiment 1 participants were free to change their decisions before they submitted their final response (a "final-choice" rule). In Experiment 2, they were informed that only their first response would be recorded (a "first-choice" rule). This was to equate the response time between choosers and nonchoosers because nonchoosers only had to click on the "Not Present" button to submit their response. Immediately after the identification decision, they were asked to make a confidence judgment with a 0 to 100 scale. At the end of Lineup 2, participants were thanked for their participation and debriefed.

3.2 Results

3.2.1 Frequency of Response Types

Table 12 shows the frequency of identification responses for target-present lineups and Table 13 shows the frequency of identification responses for target-absent lineups. Because participants were given 48 hours to complete the follow-up sessions, the length of delay was not exactly one day later or five days later. For example, the length of delay between the video and Lineup 1 was similar for both the 10mins-1day and 10mins-5days condition because Lineup 1 was presented immediately after the distractor tasks. On the other hand, the length of delay between the video and Lineup 1 varied between 5 to 7 days for the 5days-1day and 5days-5days condition due to the 48-hour window to complete the follow-up. Similarly, this 48-hour window also applied to Lineup 2. In short, because of the 48-hour window, the length of delay did vary somewhat from participant to participant.

3.2.2 The Effects of Repeated Lineups and Delay on Choosing

We were interested in the same planned comparisons as those in Experiment 1. First, we examined whether the tendency to choose in Lineup 1 and Lineup 2 differed. Second, we compared the tendency to choose across delay conditions. Unlike Experiment 1, there were four delay conditions in Experiment 2. However, the length of the initial delay (the delay between the Event and Lineup 1) was the same for the two short-initial delay conditions (10mins-1day and 10mins-5days) and also for the two long-initial delay conditions (5days-1day and 5days-5days). In other words, our comparison for the Lineup 1 choosing rates would be between the two different initial-delay lengths (10 mins vs. 5 days). Therefore, for Lineup 1, we combined the two short-initial delay conditions into a short-initial delay condition (10 mins); likewise, we combined the two long-initial delay conditions into a long-initial delay condition (5 days).

Table 12

Proportions of Identification Responses for Target-Present Lineups

		Choosers		Suspect		Lure		Not present		Total
Lineup	Delay	No.	%	No.	%	No.	%	No.	%	No.
1	10mins-1day	84	81.6	41	39.8	43	41.7	19	18.4	103
	10mins-5days	87	81.3	52	48.6	35	32.7	20	18.7	107
	5days-1day	80	72.1	29	26.1	51	45.9	31	27.9	111
	5days-5days	79	65.3	23	19.0	56	46.3	42	34.7	121
	Overall	330	74.7	145	32.8	185	41.9	112	25.3	442
2	10mins-1day	91	88.3	40	38.8	51	49.5	12	11.7	103
	10mins-5days	93	86.9	50	46.7	43	40.2	14	13.1	107
	5days-1day	90	81.1	24	21.6	66	59.5	21	18.9	111
	5days-5days	86	71.1	24	19.8	62	51.2	35	28.9	121
	Overall	360	81.4	138	31.2	222	50.2	82	18.6	442

Table 13

Proportions of Identification Responses for Target-Absent Lineups

		False Identifications		Correct Rejections		Total
Lineup	Delay	No.	%	No.	%	No.
1	10mins-1day	70	64.2	39	35.8	109
	10mins-5days	62	59.6	42	40.4	104
	5days-1day	59	54.1	50	45.9	109
	5days-5days	74	62.2	45	37.8	119
	Overall	265	60.1	176	39.9	441
2	10mins-1day	84	77.1	25	22.9	109
	10mins-5days	75	72.1	29	27.9	104
	5days-1day	79	72.5	30	27.5	109
	5days-5days	100	84.0	19	16.0	119
	Overall	338	76.6	103	23.4	441

Before merging, we examined whether the two short-initial delay conditions differed from each other in Lineup 1 choosing rates. Similarly, we compared the two long-initial delay conditions to see whether they differed in Lineup 1 choosing rates. None of these Lineup 1 comparisons reached significance for both target-present and target absent choosing rates. Thus, we merged the respective delay conditions. On the other hand, we did not merge any conditions for the Lineup 2 comparisons across delay. Although there were only two different subsequent-delay lengths (1 day vs. 5 days), the four conditions differed in their initial-delay length. Therefore, unlike Lineup 1, we examined the effect of delay on Lineup 2 choosing rates across four conditions. These same analyses were also performed for identification accuracy.

As in Experiment 1, a separate generalized mixed model was also used to analyze the choosing rates in target-present and target-absent lineups. Repeated lineups and delay conditions were the predictors; participants were specified as a random effect for both models. The status of choosing was the dependent variable (choosers were coded as 1 and nonchoosers were coded as 0). Again, because we were interested in specific comparisons, we constructed a no-intercept model and performed simultaneous tests for general linear hypotheses via the multcomp package. Adjusted p-values were reported for these multiple comparisons.

Target-Present Lineups. There was a main effect of repeated lineups, Beta = -1.74, $z = -2.50$, $p < .05$. The choosing rate tended to increase when the lineup was repeated. Next, we examined the effect of delay. There was a main effect of delay, Beta = 2.94, $z = 4.23$, $p < .001$. Our planned comparisons also indicated significant differences. For Lineup 1, choosing rate was greater in the short-initial delay condition (81.5%) than the long-initial delay condition (68.7%), Beta = 1.38, $z = 3.02$, $p < .05$. For Lineup 2, the choosing rate was significantly greater in the 10mins-1days condition (88.3%) than the 5days-5days condition (71.1%), Beta = 1.13, $z = 3.07$, *p* <.05. Similarly, the choosing rate was greater in the 10mins-5days condition (86.9%) than the 5days-5days condition (71.1%), Beta = .99, *z* = 2.84, *p* <.05. Thus, participants were more inclined to choose when lineups were repeated and were less likely to choose when the initial-delay was long. Overall, the choosing rates in the present experiment were generally higher than those in Experiment 1. This might be due to the "first-choice" rule procedure in which participants only had to click on the photo of the lineup member without having to click a "next" button to submit their response. Therefore, the results reflect the participants' first choice because they knew that they could not change their decision right before submitting their response.

Target-Absent Lineups. As in Experiment 1, choosing rate increased when the lineup was repeated. There was a main effect of repeated lineups, Beta = -3.15, *z* = -5.25, *p*<.001. We did not observed a main effect of delay. In short, there was no clear effect of delay on target-absent choosing, but repeated lineups did increase the tendency to choose across all four delay conditions. Said differently, the number of correct rejections decreased with repeated lineups.

3.2.3 The Effects of Repeated Lineups and Delay on Accuracy

The analyses for identification accuracy were the same those conducted for choosing rates (see Table 12). There was no main effect of repeated lineups. Suspect identifications did not increase with repeated lineups. In fact, most of the delay conditions showed a decrease in suspect identifications. For instance, suspect IDs decreased with repeated lineups in the 10mins-1day (39.8% to 38.8%), the 10mins-5days (48.6% to 46.7%), and the 5days-1day condition (26.1% to 21.6%). A slight increase in the suspect IDs was observed in the 5days-5days condition (19.0 to 19.8%). On the other hand, we did observed an effect of delay. For Lineup 1, suspect ID rate was greater in the short-initial delay condition (44.2%) than the long-initial delay condition (22.6%), Beta = 2.02, *z* = 4.78, *p* <.001. For Lineup 2, the suspect ID rate in the 10mins-1day condition

(38.8%) was greater than the 5days-1day condition (21.6%), Beta = .83, $z = 2.72$, $p < .05$, and the 5days-5days condition (19.8%), Beta = .94, $z = 3.09$, $p < .05$. Similarly, for Lineup 2, the suspect ID rate in the 10mins-5days condition (46.7%) was also greater than the 5days-1day condition (21.6%), Beta = 1.16, $z = 3.84$, $p < .001$, and the 5days-5days condition (19.8), Beta = 1.27, $z = 4.23$, $p < .001$. In short, the number of suspect IDs did not increase with repeated lineups, but it decreased as a function of delay.

3.2.4 Identification Decisions and Average Confidence

Once again, the confidence ratings were analyzed with linear mixed-effects regression (LMER, using the lme4 package in R). The LMER results were reported in terms of F tests. Degrees of freedom were approximated with Type II Satterthwaite approximation. All follow-up tests were performed using simultaneous tests for general linear hypotheses via the multcomp package in R.

Target-Present Lineups. Figure 10 shows the average confidence of choosers and nonchoosers in target-present Lineup 1 and Lineup 2. We conducted to a 4 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. Results indicated that there was a main effects of delay, $F(3,435.07) = 20.43$, $p < .001$. Participants in the short-initial delay conditions more confident than those in the long-initial delay conditions. There was a main effect of repeated lineups, $F(1,432.14) = 5.75$, $p < .05$. Lineup 1 confidence ratings were generally higher than Lineup 2 confidence ratings. There was no main effect of choosing status. Results showed a two-way interaction between delay and repeated lineups, $F(3,433.02) = 7.08$, $p < .001$, a two-way interaction between delay and choosing status, $F(3,736.53) = 3.96$, $p < .05$, and a two-way interaction between repeated lineups and choosing status, $F(1,489.86) = 5.49$, $p < .05$. Follow-up tests of the delay x repeated lineups interaction revealed that participants in the

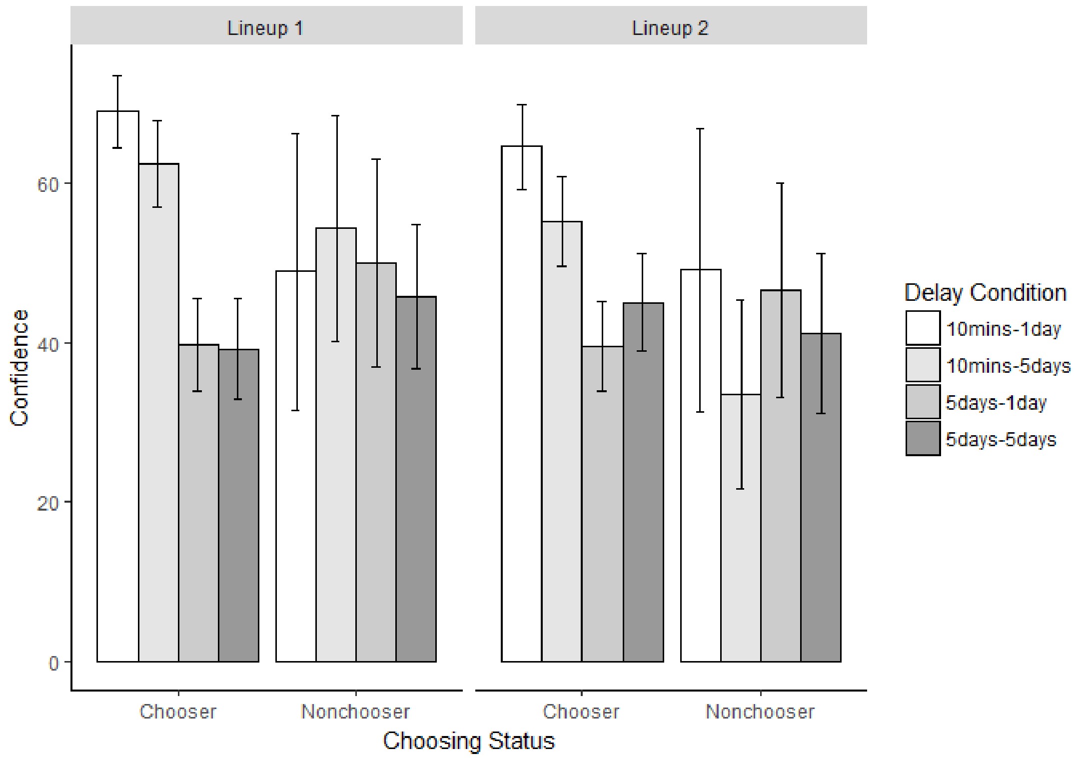


Figure 10. The average confidence of choosers vs. nonchoosers for target-present Lineup 1 and Lineup 2 in Experiment 2.

10mins-5days condition were significantly more confident in Lineup 1 than Lineup 2, $z = 2.73$, $p < .05$. In all other delay conditions, Lineup 1 and Lineup 2 confidence ratings were not significantly different. Follow-up tests of the delay x choosing status interaction revealed that choosers in 10mins-1day condition were more confident than the 5days-1day condition, $z = 9.26$, $p < .001$, and the 5days-5days condition, $z = 8.35$, $p < .001$. Similarly, choosers in 10mins-5days condition were more confident than the 5days-1day condition, $z = 6.58$, $p < .001$, and the 5days-5days condition, $z = 5.68$, $p < .001$. However, nonchoosers did not differ across delay conditions. Follow-up tests were performed for the repeated lineups x choosing status interaction. Choosers did not differ across Lineup 1 and Lineup 2. Although nonchoosers were more confident in Lineup 1 (49.0%) than Lineup 2 (42.5%), this difference did not reach significance. In sum, these results suggested that participants were generally more confident in Lineup 1 than Lineup 2; choosers in the long-initial delay conditions were less confident than those in the short-initial delay conditions.

To examine whether there were differences in the level of confidence between correct and incorrect choosers, average confidence was calculated based on each of the three possible identification responses for a target-present lineup. Figure 11 shows the average confidence of the three types of identification responses in Lineup 1 and Lineup 2. The average confidence of identification responses was submitted to a 4 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 3 (identification responses) mixed model. Results indicated that there was a main effect of delay condition, $F(3,432.27) = 17.06$, $p < .001$. Participants in the short-initial delay conditions more confident than those in the long-initial delay conditions. The main effect of repeated lineups did not reach significance, but there was a main effect of identification responses, $F(2,791.85) = 21.59$, $p < .001$. Suspect IDs were made with greater confidence than lure IDs and "Not Present"

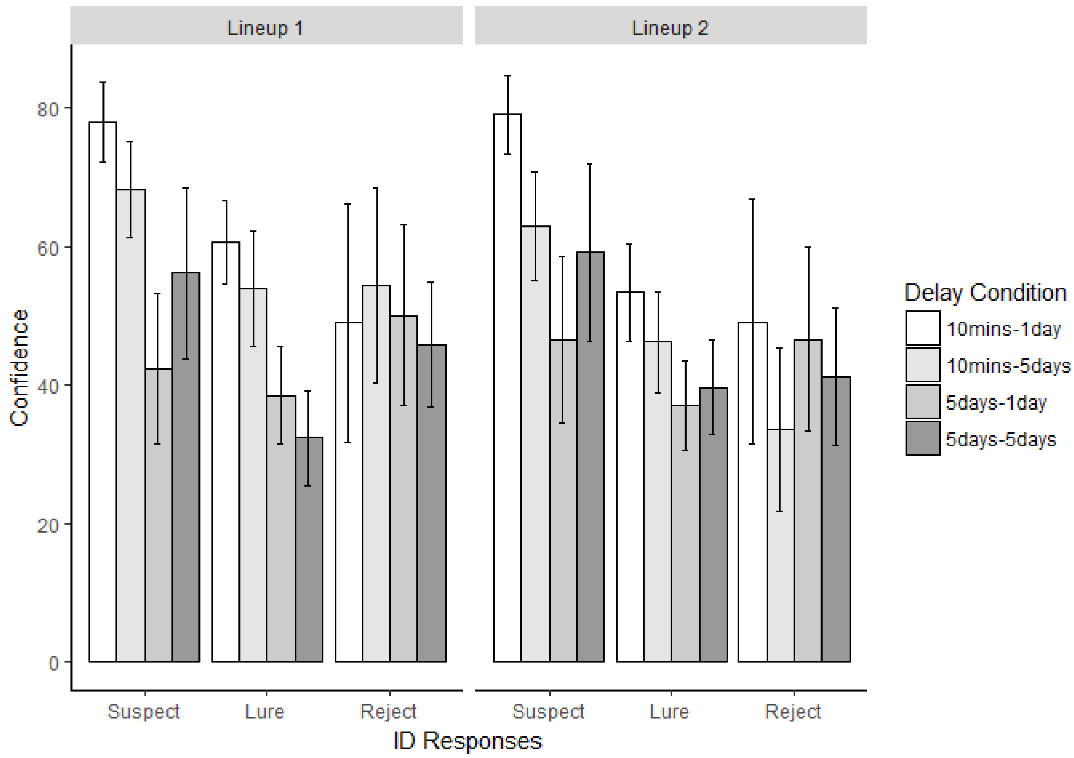


Figure 11. The average confidence of the three target-present identification responses across Lineup 1 and Lineup 2 in Experiment 2.

responses. There was also a two-way interaction between delay and repeated lineups, $F(3, 426.49) = 7.23$, $p < .001$, a two-way interaction between delay and identification responses, $F(6, 804.19) = 2.51$, $p < .05$, and a two-way interaction between repeated lineups and identification responses, $F(2,465.86) = 4.29$, $p < .05$. The three-way interaction was not significant. The follow-up tests of the delay x repeated lineups interaction showed that participants in the 10mins-5days condition were significantly more confident in Lineup 1 than Lineup 2, $z = 2.81$, $p < .05$. All other comparisons did not reach significance. Follow-up tests were also conducted on the delay x identification responses interaction. For the 10mins-1day condition, suspect IDs were made with greater confidence than lure IDs, $z = 5.39$, $p < .001$, and “Not Present” responses, $z = 5.22$, $p < .001$. For the 10mins-5days condition, suspects IDs were also made with greater confidence than lure IDs, $z = 3.92$, $p < .05$, and “Not Present” response, $z = 4.12$, $p < .001$. This pattern was not observed for the 5days-1day condition but was present in the 5days-5days condition. For the 5days-5days condition, the average confidence of suspect IDs was higher than lure IDs, $z = 4.77$, $p < .001$, and “Not Present” responses, $z = 2.91$, $p < .05$. Lastly, the follow-up tests were also conducted for the repeated lineups x identification responses interaction. For Lineup 1, suspect IDs were made with greater confidence than lure IDs, $z = 4.88$, $p < .001$, and “Not Present” responses, $z = 3.26$, $p < .05$. The same pattern was observed in Lineup 2. For Lineup 2, suspect IDs were made with greater confidence than lure IDs, $z = 6.08$, $p < .001$, and “Not Present” responses, $z = 4.94$, $p < .001$. Overall, these results suggested that correct choosers were typically more confident than incorrect choosers and those who incorrectly rejected the lineup. Participants were also generally more confident in Lineup 1 than Lineup 2; however, only the 10mins-5days condition showed a significant dip in confidence from Lineup 1 to Lineup 2.

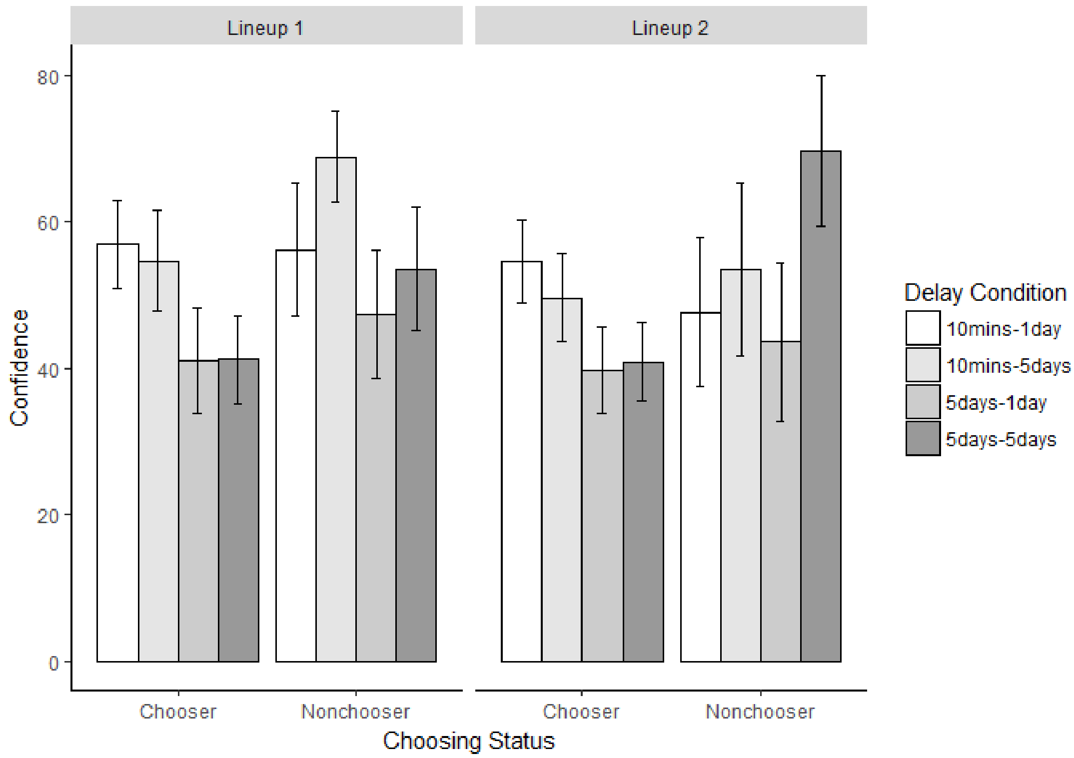


Figure 12. The average confidence of choosers vs. nonchoosers for target-absent Lineup 1 and Lineup 2 in Experiment 2.

Target-Absent Lineups. Figure 12 shows the average confidence of choosers and nonchoosers in target-absent Lineup 1 and Lineup 2, respectively. We conducted a 4 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. Results indicated significant main effects of delay, $F(3,432.87) = 8.18$, $p < .001$. Participants in the two short-initial delay conditions were more confident than those in the two long-initial delay conditions. There was a main effect of repeated lineups, $F(1,449.56) = 7.17$, $p < .05$. Participants were more confident in Lineup 1 than Lineup 2. Results showed a main effect of choosing status, $F(1,767.04) = 16.10$, $p < .001$. Nonchoosers (those who made a correct rejection) were more confident than choosers (those who made a false ID). There was also a two-way interaction between delay and repeated lineups, $F(3,449.36) = 4.29$, $p < .05$. Follow-up tests of the delay x repeated lineups interaction revealed that participants in the 10mins-5days condition were more confident in Lineup 1 than in Lineup 2, $z = 2.59$, $p < .05$. Although the other three conditions showed a decrease in confidence from Lineup 1 to Lineup 2 confidence, they did not reach significance. In short, nonchoosers (correct rejections) were more confident than choosers (false IDs). Participants generally less confident in Lineup 2 than Lineup 1; however, only the 10mins-5days condition showed a significant dip in confidence from Lineup 1 to Lineup2.

3.2.5 Confidence-Accuracy Relationship

Figure 13 shows the calibration plots for Lineup 1 and Lineup 2. The procedure for producing the calibration plots for Experiment 2 was the same as in Experiment 1. Table 14 shows the distribution of suspect and false IDs across these five confidence bins (0-20%, 21-40%, 41-60%, 61-80%, and 81-100%). As in Experiment 1, we further collapsed these five confidence bins into three confidence bins (0-40%, 41-89%, and 90-100%) to improve the stability of the calibration curves.

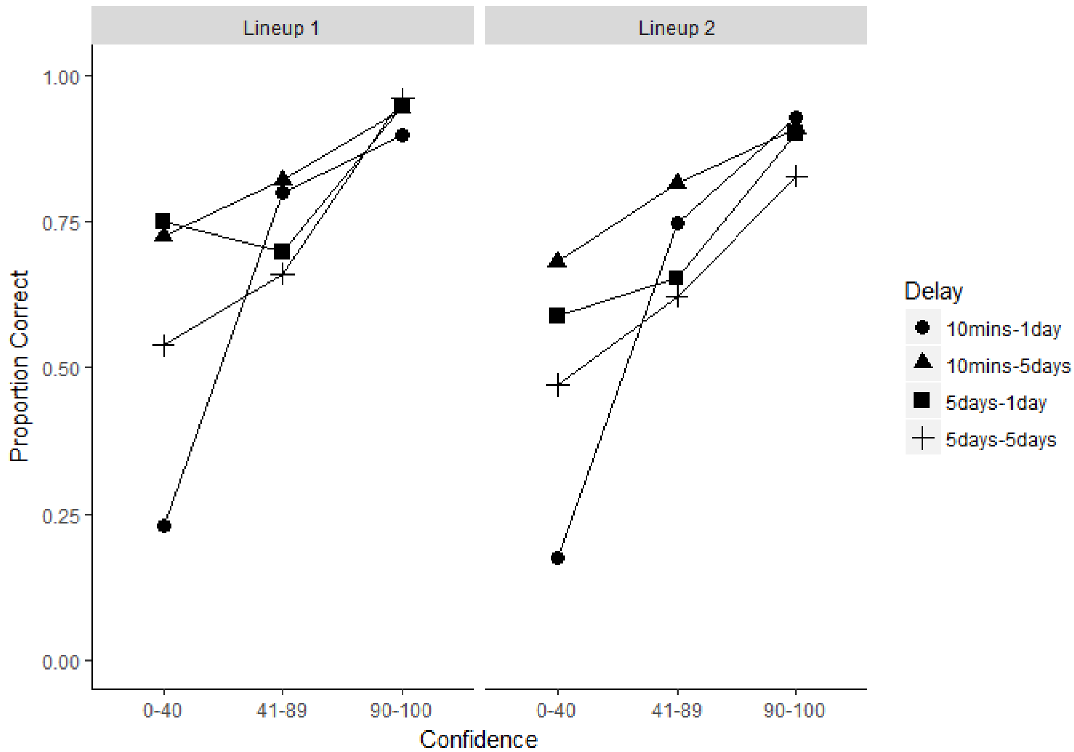


Figure 13. Experiment 2 calibration plots for choosers in Lineup 1 and Lineup 2.

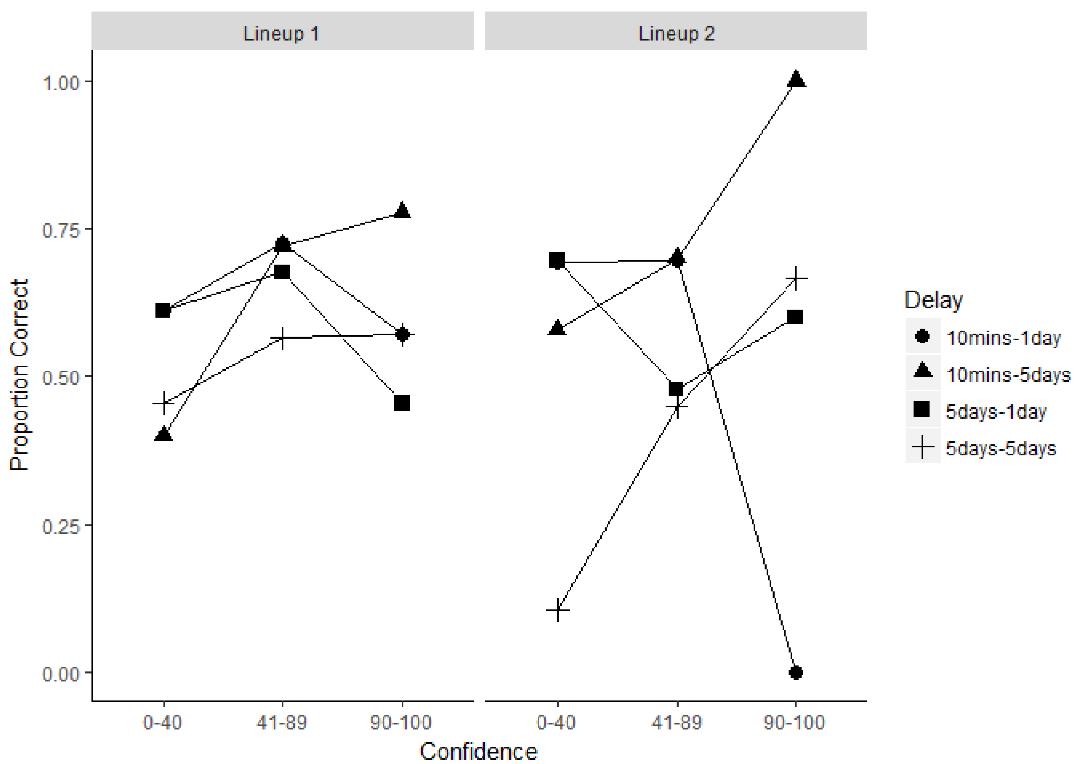


Figure 14. Experiment 2 calibration plots for nonchoosers in Lineup 1 and Lineup 2.

Table 14

The Number of Suspect and False IDs across Confidence Bins

		Lineup 1 Confidence Bins					Lineup 2 Confidence Bins				
Delay	Responses	0-20	21-40	41-60	61-80	81-100	0-20	21-40	41-60	61-80	81-100
10mins-1day	Suspect IDs	0	1	9	8	23	0	1	6	12	21
	False IDs	7	13	16	22	12	12	16	16	25	15
10mins-5days	Suspect IDs	2	5	9	17	19	6	5	12	12	15
	False IDs	10	6	14	25	7	11	20	20	16	8
5days-1day	Suspect IDs	9	7	6	4	3	7	3	7	4	3
	False IDs	19	13	8	14	5	24	18	18	15	4
5days-5days	Suspect IDs	4	3	8	4	4	4	4	3	8	5
	False IDs	16	20	19	13	6	29	25	21	16	9

The positive relationship between confidence and accuracy was observed in all the four delay conditions across Lineup 1 and Lineup 2. In all four delay conditions, the highest confidence bin (90-100%) corresponded to the highest level of accuracy. For Lineup 1, the accuracy of the highest confidence bin ranged from 90% to 96%. For Lineup 2, the accuracy of the highest confidence bin ranged from 83% to 93%. The positive confidence-accuracy relationship was observed in both Lineup 1 and Lineup 2, although the accuracy of the highest confidence bin in Lineup 1 was generally greater than Lineup 2. In short, the confidence-accuracy relationship for choosers remained intact despite the effects of delay and repeated lineups.

Again, we also constructed calibration plots for nonchoosers. Figure 14 shows the nonchoosers calibration plots for Lineup 1 and Lineup 2. The calibration curves appeared more stable in Lineup 1 than Lineup 2. However, the calibration curves for nonchoosers were still less meaningful compared to choosers.

3.2.6 Confidence Ratings and Response Times as Indicators of Identification Accuracy

As in Experiment 1, we evaluated the usefulness of both confidence judgments and response times as indicators of accuracy. Again, the calculation of accuracy was the same as the calculation for calibration [correct suspect IDs / (correct suspects IDs + incorrect suspect IDs)]. Because there was no designated innocent suspect in the target-absent lineups, the sum of incorrect suspect IDs for a confidence bin (rating of 90% or above) or a RT bin (10 seconds or earlier) was divided by the number of lineup members. Tables 15 and 16 show the accuracy for high confidence and fast response time, respectively. Again, we used the same criteria as in Experiment 1: high confidence (90% or above) or fast response time (10 seconds or earlier). We found that high confidence generally produced higher accuracy than fast RTs. In most cases, a

Table 15

Accuracy of Responses Above or Below 90% Confidence

Lineup	Delay	Confidence	% correct
Lineup 1	10mins-1day	Above	89.66
		Below	73.36
	10mins-5days	Above	94.29
		Below	80.92
	5days-1day	Above	94.74
		Below	72.90
	5days-5days	Above	96.00
		Below	60.96
Lineup 2	10mins-1day	Above	92.78
		Below	66.08
	10mins-5days	Above	90.91
		Below	77.67
	5days-1day	Above	90.00
		Below	62.07
	5days-5days	Above	82.76
		Below	55.81

Table 16

Accuracy of Responses Before or After 10 seconds

Lineup	Delay	Response Time	% correct
Lineup 1	10mins-1day	Before	81.36
		After	67.50
	10mins-5days	Before	86.36
		After	76.36
	5days-1day	Before	71.29
		After	77.27
	5days-5days	Before	59.41
		After	70.27
Lineup 2	10mins-1day	Before	78.17
		After	45.00
	10mins-5days	Before	82.17
		After	68.85
	5days-1day	Before	66.67
		After	57.69
	5days-5days	Before	59.02
		After	59.02

Table 17

Accuracy of Responses with a Confidence Rating 90% or Above, and a Response Time of 10 Seconds or Earlier

Lineup	Delay	% correct
Lineup 1	10mins-1day	89.19
	10mins-5days	83.08
	5days-1day	80.00
	5days-5days	85.71
Lineup 2	10mins-1day	94.74
	10mins-5days	91.53
	5days-1day	90.00
	5days-5days	82.76

high confidence criterion was superior to the combination of high confidence and fast RTs (see Table 17). Likewise, the accuracy of the combination of high confidence and fast RTs was generally better than just the fast RTs criterion.

3.2.7 Consistent and Inconsistent Decisions across Lineup 1 and 2

Table 18 shows the proportion of same and different decisions across Lineup 1 and 2. In both target-present and target-absent lineups, participants generally repeated their Lineup 1 decisions in Lineup 2. Although our chi-square analyses of these differences across delay conditions did not reach significance, the following patterns were consistently observed in both target-present and target-absent lineups. For example, there were generally more same decisions than different decisions when the delay between Lineups 1 and 2 was short (e.g. the 10mins-1day and the 5days-1day conditions). This also means that there were more different decisions than same decisions when the delay between Lineups 1 and 2 was long (e.g. the 10mins-5days and the 5days-5days conditions). Furthermore, the effect of the subsequent delay was also conditional on the initial delay. For instance, the proportion of same decisions was higher in the 10mins-1day condition than the 5days-1day condition. Likewise, the proportion of different decisions was higher in the 10mins-5days condition than the 5days-5days condition. In short, although we consistently observed the hypothesized patterns in both target-present and target-absent lineups, these effects were not significant.

Similar to the results in Experiment 1, confidence ratings decreased with repeated lineups regardless whether people made the same or a different decision (see Table 19). There was only one condition that did not conform to this pattern. The different decisions in the target-present 5days-5days condition actually showed an increase in confidence at Lineup 2. Their Lineup 1 confidence was 37.2% and their Lineup 2 confidence was 42.9%. Moreover, the average

Table 18

Proportion of Same and Different Decisions across Lineup 1 and 2

Lineup Type	Delay	Same Decision	Different Decision
Target-Present	10mins-1day (N = 103)	70.9	29.1
	10mins-5days (N = 107)	66.4	33.6
	5days-1day (N = 111)	60.4	39.6
	5days-5days (N = 121)	53.7	46.3
Target-Absent	10mins-1day (N = 109)	63.3	36.7
	10mins-5days (N = 104)	54.8	45.2
	5days-1day (N = 109)	61.5	38.5
	5days-5days (N = 119)	54.6	45.4

Table 19

Confidence Ratings of Same and Different Decisions across Lineup 1 and 2

		Same		Different	
Lineup Type	Delay	Lineup 1	Lineup 2	Lineup 1	Lineup2
Target-Present	10mins-1day	70.4	67.6	53.1	51.3
	10mins-5days	63.7	55.7	55.7	45.9
	5days-1day	41.3	40.6	44.8	41.3
	5days-5days	45.2	45.0	37.2	42.9
Target-Absent	10mins-1day	59.3	56.1	52.1	47.8
	10mins-5days	63.3	53.9	57.0	46.9
	5days-1day	47.5	43.6	38.2	36.4
	5days-5days	48.9	50.0	42.3	40.1

confidence of the two short-initial delay conditions (e.g. the 10mins-1day and the 10mins-5days conditions) was generally higher than the two long-initial conditions (e.g. the 5days-1day and the 5days-5days conditions). This was true for both target-present and target-absent lineups, irrespective of whether it was same or different decisions. In short, although participants generally made the same decision in both Lineups 1 and 2, the length of delay between Lineup 1 and Lineup 2 appeared to affect the likelihood of doing so. Like the results of Experiment 1, participants were typically less confident in their Lineup 2 decision than their Lineup 1 decision regardless of whether they repeated a decision or made a different decision.

3.2.8 Lineup 1 and 2 Decision Combinations and Their Confidence Ratings

As in Experiment 1, we also divided the same and different decisions into specific subgroups of interest. Tables 20 and 21 show the proportion and confidence ratings of the different decision combinations for both target-present and target-absent lineups, respectively.

Target-Present Lineups. Across all four delay conditions, there was a larger proportion of repeated suspect IDs compared to the other two target-present decision combinations (repeated Lure IDs and repeated "Not Present" responses). A sizeable proportion of repeated Lure IDs and repeated "Not Present" responses was also observed. Consistent with the results of Experiment 1, participants who made an incorrect decision in Lineup 1 may continue to do so in Lineup 2. The average confidence of these decision combinations were generally lower in Lineup 2 than Lineup 1. Some of the decision combinations in the two long-initial delay conditions did show an increase in confidence. For example, repeated suspect IDs for the 5days-1day condition started with an average confidence rating of 41.9% in Lineup 1 and it increased to 43.6% in Lineup 2. Similarly, the average confidence of repeated suspect IDs for the 5days-

Table 20

The Count and Confidence Ratings of Decision Types across Target-Present Lineup 1 and 2

			No. of Decisions			Confidence Ratings	
Lineup Type	Delay	Response Type	Lineup 1	Lineup 2	%	Lineup 1	Lineup 2
Target-Present	10mins-1day	Repeated Suspect ID	41	34	82.9	81.2	81.4
		Repeated Lure ID	43	32	74.4	61.7	58.2
		Repeated "Not Present"	19	7	36.8	57.4	43.3
		"Not Present" to Choosing	19	12	63.2	44.0	43.8
	10mins-5days	Repeated Suspect ID	52	45	86.5	68.6	65.0
		Repeated Lure ID	35	17	48.6	55.4	40.2
		Repeated "Not Present"	20	9	45.0	54.8	38.4
		"Not Present" to Choosing	20	11	55.0	54.1	37.6
	5days-1day	Repeated Suspect ID	29	20	69.0	41.9	43.6
		Repeated Lure ID	51	32	62.7	36.7	36.2
		Repeated "Not Present"	31	15	48.4	36.3	32.4
		"Not Present" to Choosing	31	16	51.6	49.5	43.2
	5days-5days	Repeated Suspect ID	23	15	65.2	60.5	62.6
		Repeated Lure ID	56	28	50.0	35.5	38.4
		Repeated "Not Present"	42	22	52.4	47.1	41.1
		"Not Present" to Choosing	42	20	47.6	44.2	44.9

Table 21

The Count and Confidence Ratings of Decision Types across Target-Absent Lineup 1 and 2

			No. of Decisions			Confidence Ratings	
Lineup Type	Delay	Response Type	Lineup 1	Lineup 2	%	Lineup 1	Lineup 2
Target-Absent	10mins-1day	Repeated Lure ID	70	41	58.6	59.9	58.7
		Repeated "Not Present"	39	19	48.7	57.5	47.2
		"Not Present" to Choosing	39	20	51.3	55.0	46.0
	10mins-5days	Repeated Lure ID	62	33	53.2	59.1	52.1
		Repeated "Not Present"	42	19	45.2	72.4	54.6
		"Not Present" to Choosing	42	23	54.8	66.0	51.0
	5days-1day	Repeated Lure ID	59	34	57.6	46.2	44.7
		Repeated "Not Present"	50	26	52.0	51.5	43.2
		"Not Present" to Choosing	50	24	48.0	42.9	37.8
	5days-5days	Repeated Lure ID	74	40	54.1	40.2	42.3
		Repeated "Not Present"	45	16	35.6	68.0	68.4
		"Not Present" to Choosing	45	29	64.4	45.7	39.1

5days condition was 60.5% in Lineup 1 and it increased to 62.6% in Lineup 2. However, none of these exceptions showed an increase in confidence exceeding 5%.

Target-Absent Lineups. In all four delay conditions, lure IDs in Lineup 1 were often repeated in Lineup 2. 41 of 70 lure IDs (58.6%) were repeated in the 10mins-1day condition and 33 of 62 lure IDs (53.2%) were repeated in the 5days-1day condition. Repeated lure IDs were also observed in the two long-initial delay conditions. 34 of 59 lure IDs (57.6%) were repeated in the 5days-1day condition and 40 of 74 lure IDs (54.1%) were repeated in the 5days-5days condition. Although these proportions were only slightly more than 50%, they showed that there was a tendency to repeat an error (a lure ID in a target-absent lineup). There were participants who repeated their correct rejection across Lineup 1 and 2; however, there was also a considerable number of participants who had initially correctly rejected the lineup but chose someone at the subsequent lineup. Again, participants were typically less confident in their Lineup 2 decisions than their Lineup 1 decisions. In short, it appeared that participants were generally inclined to repeat their Lineup 1 decisions (a correct or incorrect decision) while others may switch from a correct decision to an incorrect one. It did not matter which of the three types of target-absent decision combinations, the average confidence of Lineup 1 was generally higher than Lineup 2.

Overall, these results were consistent with Experiment 1. Experiment 2 did show that the length of delay between Lineup 1 and 2 has the potential to affect the likelihood of whether someone repeats their Lineup 1 decisions. Participants were generally less confident about their Lineup 2 decisions than Lineup 1 decisions, irrespective of whether they were repeating a correct or incorrect decision or switching form a correct decision into an incorrect one.

Table 22

The Proportion of Consistent Responses Above and Below each Confidence Level

Delay	Confidence	Above	Below
10mins-1day	0	67.0	NA
	20	69.6	42.9
	40	71.5	47.5
	60	75.4	53.7
	80	**77.2**	63.2
	100	**72.7**	66.7

Delay	Confidence	Above	Below
5days-1day	0	60.9	NA
	20	63.0	55.2
	40	62.3	59.4
	60	65.8	58.3
	80	**65.6**	60.1
	100	**75.0**	60.7

Delay	Confidence	Above	Below
10mins-5days	0	60.7	NA
	20	61.7	42.9
	40	62.8	47.5
	60	64.5	53.7
	80	**72.2**	63.2
	100	**77.8**	66.7

Delay	Confidence	Above	Below
5days-5days	0	54.2	NA
	20	56.5	46.4
	40	56.7	50.9
	60	63.6	49.7
	80	**73.3**	51.4
	100	**83.3**	53.4

3.2.9 Confidence and Decision Consistency

As in Experiment 1, we also examined whether the likelihood of repeating a decision was related to the Lineup 1 confidence rating. Again, we calculated the proportion of consistent decisions above and beyond the median and by each confidence level (20%, 40%, 60%, 80%, and 100%).

Confidence Levels. Table 22 shows the proportion of consistent responses above and below each confidence level. Consistent with the results of Experiment 1, decisions were more likely to be repeated if they were made with high confidence in Lineup 1. This relationship was present in all four delay conditions, especially at the two upper confidence levels (at 80% and 100%). Although this relationship was not perfect, Lineup 1 decisions made with high confidence were generally likely to be repeated regardless of delay.

3.3 Discussion

Overall, the choosing rates in the Experiment 2 were greater than those in Experiment 1; however, we observed similar patterns of results. Similar to the findings in Experiment 1, target-present choosing rate decreased as a function of delay. On the other hand, there was not a clear pattern for target-absent choosing rate. Regardless of the length of delay or the lineup type (target-present or target-absent), participants were more inclined to choose when the lineup was repeated. Similarly, suspect ID rate decreased as a function of delay, and repeated lineups did not lead to an increase in suspect IDs.

There were some variabilities in choosing rate and suspect ID rate between delay conditions with similar initial-delay lengths. For instance, although the length of initial delay was the same for the 10mins-1day and the 10mins-5days condition, the proportion of suspect IDs made at Lineup 1 differed (39.8% vs. 48.6%). Their target-present choosing rates were similar

(81.6% and 81.3%, respectively). The dissimilar Lineup 1 choosing rates and suspect ID rates in the two long-initial delay conditions (5days-1day and 5days-5days) were less surprising because these participants had a 48-hour window to complete the follow-up sessions after the initial 5-day delay. However, the different follow-up completion times did not produce a significant difference in delay conditions with the same initial-delay length. That is, for Lineup 1, the two short-initial delay conditions (10mins-1day and the 10mins-5days) were not significant different in choosing rate or suspect ID rate; likewise, this was also true for the two long-initial delay conditions (5days-1day and 5days-5days).

Moreover, despite the variability, these results were consistent with our predictions. The Lineup 1 target-present choosing rates and suspect ID rates in the two initial-long delay conditions were lower than those in two initial short-delay conditions. Furthermore, the suspect IDs results were also consistent with our prediction that repeated lineups should maintain or decrease the proportion of suspect IDs. As hypothesized, there was no significant difference in suspect IDs between Lineup 1 and Lineup 2. In sum, target-present choosing rates and suspect IDs decreased as a function of delay. Regardless of lineup type, repeated lineups increased the tendency to choose, but this increase in choosing did not lead to an increase in the number of suspect IDs in Lineup 2.

In general, correct decisions were made with higher confidence than incorrect decisions. For target-present lineups, correct choosers were more confident than incorrect choosers. For target-absent lineups, nonchoosers (correct rejections) were generally more confident than choosers (false IDs). Furthermore, participants in the two short-initial delay conditions were generally more confident than those in two long-initial delay conditions. They were also usually more confident in Lineup 1 than Lineup 2. For both target-present and target-absent lineups, the

10mins-5days condition showed a significant decrease in confidence at Lineup 2. One possible explanation is the length of the initial delay and the subsequent in this particular condition. People were generally more confident with a short-initial delay than a long-initial delay. However, when the subsequent delay was long, they became less confident. Perhaps this pattern was not observed in the 5days-5days condition because people in the long-initial delay conditions were less confident in Lineup 1 and their confidence leveled off at Lineup 2. This appears to be the case. In Lineup 1, participants in the 10mins-1day and 10mins-5days condition were generally more confident than those in the 5days-1day and 5days-5days condition. Compared to the 10mins-5days condition, those in the 5days-1day and the 5days-5days condition had a similar level of confidence across Lineup 1 and Lineup 2. While both the 10mins-1day and 10mins-5days condition showed a dip in confidence at Lineup 2, the 10mins-5days condition showed a greater decrease in Lineup 2 confidence. In sum, both the effects of delay (the initial delay and the subsequent delay) and repeated lineups can influence the level of confidence expressed.

Even though the effects of repeated lineups and delay influenced confidence judgments, there was a positive relationship between confidence and accuracy for choosers. High confidence continued to be associated with high accuracy and it remained so even when across delay conditions and repeated lineups, which demonstrated the robustness of the confidence-accuracy relationship. On the other hand, the confidence-accuracy relationship for nonchoosers appeared less reliable.

We also showed that confidence was superior to response time as an indicator of identification accuracy. Identification accuracy was higher using a high confidence criterion (90% or above) than a fast RT criterion (10 seconds or earlier). The combination of high confidence and fast RT was also superior to a fast RT criterion. Moreover, we found the optimal time-boundary varied across different manipulations (see Appendix B). In other words, there was no specific RT cutoff that best differentiated correct decisions from incorrect decisions across different delay lengths and repeated lineups. In short, confidence judgments were more reliable than RTs.

In addition, we were also interested in the proportion of consistent and inconsistent responses as a function of the length of the initial and subsequent delay. This was the primary goal of Experiment 2. We expected participants to generally repeat their Lineup 1 decisions in Lineup 2; however, we hypothesized that the length of the subsequent delay influenced the likelihood of repeated decisions. For instance, they would be more likely to repeat a decision when the subsequent delay was short than when it was long. Although our chi-square analyses were not significant, the hypothesized patterns of results were consistently observed in both target-present and target-absent lineups. There was a larger proportion of same decisions in the 10mins-1day condition than the 10mins-5days condition. Likewise, there were more same decisions in the 5days-1day condition than the 5days-5days condition. By extension, these results also suggest that a long subsequent delay increases the likelihood of different decisions compared to a short subsequent delay. This effect of the subsequent delay was also conditional on the length of the initial delay. That is, if the length of the subsequent delay was the same, then the proportion of same decisions should be larger for a condition with a short initial delay than a condition with a long initial delay. This appears to be the case. For example, the 10mins-1day condition had a larger proportion of same decisions than the 5days-1day condition; however, this pattern was more apparent in target-present lineups than target-absent lineups.

Again, repeated lineups did not lead to confidence inflation. When participants made the same decisions across Lineup 1 and 2, they typically became slightly less confident at Lineup 2 regardless of the types of lineup (target-present or target-absent), delay or identification decisions. On average, confidence ratings in Lineup 2 were lower than Lineup 1. This was true irrespective to whether participants kept the same response or made a different response in Lineup 2, which showed that people were overall less confident in their decisions at Lineup 2. Therefore, it suggests that people are likely to become less confident in the subsequent identification if they were not informed whether their Lineup 1 decision was correct (i.e. no feedback).

As in Experiment 1, we also attempted to use Lineup 1 confidence judgments to predict the likelihood of repeating a Lineup 1 decision at Lineup 2. Again, the confidence ratings were divided by confidence levels (20%, 40%, 60%, 80%, and 100%). Our results were consistent with Experiment 1. People were more likely to make the same decision again if they had initially made that decision with high confidence. This finding further confirms Korat (2012)'s claim that confidence is also a predictor of reproducibility.

Consistent with Experiment 1, the effects of repeated lineups and delay play a role in identification performance and post-dictive indicators. The unique finding of Experiment 2, however, is the effects of the subsequent delay (the delay between Lineups 1 and 2). Our results showed that the length of delay between two lineups (the subsequent delay) can influence the witnesses' tendency to stay consistent. For instance, if witnesses receive a second lineup shortly after the first lineup, they will more likely make the same identification decision again, regardless of whether the initial decision was correct or incorrect. In other words, consistency is not necessarily a confirmation of accuracy. Lastly, the most important finding is that the utility of confidence judgments. Even under various conditions (e.g., different lengths of delay or repeated lineups), the confidence-accuracy relationship remained predictive of accuracy.

Chapter 4: General Discussion

The goals of the present study were to examine the effects of repeated lineups and delay on witness decision behaviors (e.g. the tendency to choose and make the same response across lineups), accuracy, the confidence-accuracy relationship, and the reliability of common post-dictive indicators (e.g. confidence judgments and response time). While most of results were consistent across Experiments 1 and 2, a few findings were not. We will discuss each of the findings in the following sections.

4.1 Choosing

Across Experiments 1 and 2, repeated lineups consistently increased the tendency to choose from a lineup, regardless of whether the target was present or not. On the other hand, the effects of delay on choosing differed across the two experiments. For target-present lineups, both experiments consistently showed that a decrease in choosing as a function of delay on Lineup 1; this was not the case for target-absent choosing rates. Participants in Experiment 1 were less inclined to choose when the initial delay was short (the 10-min delay condition) than when it was long (the 3-day delay condition). In Experiment 2, participants were more willing to choose when the delay was short (the 10mins-1day and the 10mins-5days condition), but some participants in the long-initial delay conditions were similarly likely to choose (the 5days-5days condition) while others were less inclined to do so (the 5days-1day condition). However, the effect of delay on target-absent choosing rate did not reach significance in Experiments 1 or 2.

It is possible that target-absent choosing rate might be more variable than target-present choosing rate. Unlike the target-absent choosing rates, target-present choosing rates may be more predictable because it is partially driven by the presence of the target and people's ability to choose the suspect. People are usually able to choose the suspect at a short delay compared to a long delay. Because target-present choosing rates are the combination of suspect and lure IDs, these correct choosers contribute to the high choosing rate at a short delay and the low choosing rate at a long delay. On the other hand, target-absent choosing rate is driven solely by lure IDs. Because lure IDs are often made with low confidence, target-absent choosing rate may be a reflection of variabilities in individual characteristics in accuracy motivation and metacognitive ability. For example, people can have different thresholds for choosing. Some participants may be willing to choose someone despite having only a confidence level of 60%, whereas other participants may not choose someone unless they had a confidence level of 80%. Of course, the level of confidence also depends on their metacognitive ability. Some people are generally overconfident while others tend to be underconfident (e.g., their belief about their memory ability). In both experiments, lure IDs were generally made with lower confidence than suspect IDs, which potentially suggest that these incorrect choosers were willing to risk choosing someone despite having low confidence in their decisions. Future studies should explore individual differences in choosing behaviors.

4.2 Accuracy

Unsurprisingly, the number of suspect IDs decreased as a function of delay. This finding was replicated in both Experiments 1 and 2. Although it appeared that repeated lineups led to an increase in suspect IDs in Experiment 1, this increase was not significant. This pattern was not observed in Experiment 2. As mentioned previously, the majority of the Experiment 1 participants who made a suspect ID at Lineup 2 had initially made a “Not Present” present rather than a lure ID. Therefore, it is quite possible that these participants were highly cautious at Lineup 1 but were more willing to choose someone at Lineup 2. Steblay et al. (2013) reported similar findings and suggested that participants may apply a “close enough” judgment at Lineup

2. Of course, some of these participants could have also picked someone in Lineup 2 due to misplaced familiarity. Similar to Experiment 1, Steblay et al. (2013) also found these results with a final-choice rule. That is, participants were free to change their decisions before they submit their final response. Therefore, participants in Experiment 1 may initially select the target, but they hesitated before clicking on the “Next” button and chose the “Not Present” response instead. Based on this reasoning, removing the final-choice rule should lead to a higher Lineup 1 suspect ID rate and no increase in Lineup 2. Experiment 1 employed the final-choice rule (for choosers), whereas Experiment 2 did not. Overall, the Lineup 1 suspect ID rates in the two short-initial delay conditions (with a 10-min delay) in Experiment 2 were higher than the 10-min delay condition in Experiment 1. In addition, no obvious increase in suspect IDs was observed across repeated lineups in Experiment 2. Therefore, it is possible that the final-choice rule might be responsible for the non-significant increase in suspect IDs across repeated lineups. Because we did not provide a post-experiment questionnaire that asked participants why they made their identification decisions, we cannot confirm or eliminate the aforementioned explanations. Nevertheless, further research should examine the cause of this pattern of results and whether the “final-choice” rule vs. the “first-choice” rule produce differences in identification performance.

4.3 Decision Consistency

Over 50% of the participants made the same decisions across Lineup 1 and Lineup 2, regardless of whether the initial decision was correct or not. This pattern was observed in both Experiments 1 and 2. However, a few factors seem to affect the likelihood of repeated decisions. For example, the length of the subsequent delay (the delay between Lineup 1 and Lineup 2) was responsible for producing different proportions of same and different decisions. A short-subsequent delay led to more repeated decisions, whereas a long-subsequent delay produced more inconsistent decisions across repeated lineups. This effect was also conditional on the length of the initial delay. The proportion of repeated decisions was greater when the initial delay was short than when it was long. Furthermore, Lineup 1 confidence was also a predictor of whether a particular decision would be repeated. Compared to low confident decisions, high confident decisions in Lineup 1 were likely to be repeated in Lineup 2. Consistent with Koriat (2012)'s idea that confidence is also a predictor of reproducibility, Lineup 1 confidence was an indicator of the likelihood of repeated decisions in Lineup 2. In sum, the length of the initial- and subsequent-delay, and the level of confidence expressed at Lineup 1 can affect the likelihood of repeated decisions.

4.4 Confidence-Accuracy Relationship

Using the calibration approach, numerous eyewitness studies (Palmer et al., 2013; Sauer et al., 2010; Wixted et al., 2015; Wixted & Wells, 2017) have consistently found a positive confidence-accuracy relationship for choosers. For example, regardless of varying length of delay, exposure duration, and divided attention, Palmer et al. (2013) found a positive relationship between confidence and accuracy. In addition to these factors, the present study also found that the confidence-accuracy relationship remained intact in a repeated lineup (when no feedback provided during the initial identification). Moreover, even the combination of repeated lineups and delay was not enough to eliminate the positive confidence-accuracy relationship[1], which further shows the robustness of this relationship. In addition to biased lineups, one known exception to these findings is when it involved child witnesses (Keast, Brewer, & Wells, 2007).

Future studies should continue to test the boundary of the confidence-accuracy relationship and identify any exceptions.

On the other hand, the confidence-accuracy relationship for nonchoosers was not meaningful. This finding is consistent with prior literature (Keast et al., 2007; Palmer et al., 2013). A possible reason for the lack of a relationship for nonchoosers could be due to the how the “Not Present” response was used. Although witnesses are supposed to reject the lineup if they believe the culprit is not in the lineup, they may also use this option if they are not confident enough to choose, or they simply do not know. Weber and Perfect (2012) found that people do not usually give a “Don’t Know” response unless this option was made explicit to them. Compared to the when the “Don’t’ know” option was not available (i.e. force-report condition), the condition with the explicit option to use a “Don’t Know” response (i.e. free-report condition) had significantly better identification performance. In other words, the “Don’t Know” option screened out errors, which in turn improve the overall performance. However, their study did not examine whether the “Don’t Know” option produced changes to the calibration curve for choosers and nonchoosers. Future research should investigate the screening effect of a “Don’t Know” on the confidence-accuracy relationship.

4.5 Confidence Judgments versus Response Times

The present experiments showed that the level of confidence and response times were associated with accuracy. For example, correct choosers were, on average, more confident and made their decisions faster than incorrect choosers (see Appendix A). The patterns of results for nonchoosers were generally mixed. When a lineup was repeated, people generally less confident in Lineup 2 than Lineup 1 across; they also made their subsequent lineup decisions faster.

Interestingly, there was a delay effect on the level of confidence for choosers. Choosers in a short-delay condition were generally more confident than those in the long-delay condition. This suggests that people were aware of their poorer memory after a long retention and adjusted their confidence level accordingly. The higher level of confidence was also observed for choosers in target-absent lineups, which indicates that the high confidence here was a reflection of the delay effect rather than accuracy. These incorrect choosers were generally less confident than correct choosers. In other words, high confidence was associated with correct identification (i.e. suspect IDs).

Compared to confidence judgments, response time was less telling. Consistent with prior studies (Dunning & Perretta, 2002; Smith et al., 2000), correct choosers were faster than incorrect choosers; however, there were more variability in the response time results. This can be explained by the nature of these two measures. While confidence ratings are bounded by the scale (e.g. 0 to 100), response time does not have an upper bound. It is less surprising to observe more variability in response times than in confidence ratings. Therefore, in addition to averages, we evaluated confidence judgments and response times with criteria used in other studies (Dunning & Perretta, 2002; Weber et al., 2004).

We examined the accuracy associated with a high confidence criterion (90%) or a fast response time criterion (10 sec or earlier). Across both experiments, we consistently found that a high confidence criterion outperformed the fast response time criterion. In contrast to Weber et al. (2004), the present study showed that the combination of post-dictive indicators was not superior to either indicator alone. The combination of both high confidence and fast response time was superior to a fast response time criterion, but it was no better than a high confidence criterion. One possible explanation for this is that our response time data is more variable. Unlike prior response time studies (Brewer et al., 2006; Dunning & Perretta, 2002; Weber et al., 2004), the present experiments were online studies where the participants could be inattentive or distracted. Furthermore, our sample consisted of a wide age range, whereas prior studies had undergraduate participants. All these factors could have contributed to the variability in our data.

Aside from these factors, other studies have also found variable response time results. For example, Weber et al. (2004) found the optimal time-boundary that differentiate correct choosers from incorrect choosers differed across all four of their studies. Likewise, our experiments also found inconsistent time-boundaries across all of our conditions (see Appendix B). Because of inconsistencies like these, it is difficult to establish a reliable response time cutoff that can differentiate correct identification decisions from incorrect ones. This is not to say that response time is an unusable indicator of accuracy, but it is more susceptible to variability compared to a bounded scale like confidence judgments. Future research should examine factors that help response time become a more reliable measure. Nonetheless, given our current findings, confidence appears to be the more reliable index of accuracy.

In sum, the present study showed that the effects of repeated lineups and delay play a role in witness decision behaviors (e.g. the tendency to choose), identification accuracy, and post-dictive indicators of accuracy (confidence and response time). Even under both the effects of delay and repeated lineups, we also found the confidence-accuracy relationship remained predictive of accuracy. In addition, we showed that confidence is also predictor of reproducibility. Lastly, confidence is a more reliable indicator of accuracy than response time.

4.6 Limitations

Compared to studies done in lab settings, we had less control in our studies because we used online participants. Aside from distracted participants, we also did not have control of the presentation of the study materials. For example, varying monitor size and brightness setting could have affected the quality of the video and the lineup photos. We also could not control the environment in which these participants completed the experiment. Furthermore, our sample consisted of individuals with a wide age range and possibly different level of cognitive abilities compared to the typical undergraduate students used in prior eyewitness studies. Despite of these factors, the present study consistently found a positive confidence-accuracy relationship. This consistent finding provides further evidence that the confidence-accuracy relationship is robust.

4.7 Implications

Based on our comprehensive analysis of the effects of repeated lineups and delay on eyewitness identification, we showed that the impact of these influences can be observed at a basic level (e.g. changes in decision behaviors such the tendency to choose from a lineup) and a higher level (e.g. the confidence-accuracy relationship). Together, the different levels of analysis provide a more complete picture of these effects. Given these findings, we showed that these factors (repeated lineups and delay) should be taken into consideration when conducting eyewitness identification and evaluating reliability of a given identification decision.

Repeated Lineups. Although the present study showed that the confidence-accuracy remained intact even at a repeated lineup, should we use repeated lineups? We consistently found that people became more willing to choose when they received a subsequent lineup. These subsequent identifications often consisted of innocent filler IDs, which may suggest misplaced familiarity. In other words, people were making more errors in the subsequent lineup because the faces in the subsequent lineup might appear familiar to them after seeing them in the initial lineup. We also showed incorrect decisions in the initial lineup were not likely to be corrected in the subsequent lineup. For example, a majority of the participants repeated their Lineup 1 decisions in Lineup 2, regardless of whether their Lineup 1 decisions were correct or incorrect. In short, witnesses may appear consistent (or reliable) when they make the same decision again, but consistency should not be taken as an indication or confirmation of accuracy.

In sum, the effects of repeated lineups and delay can influence witness decision behaviors (e.g. the tendency to choose and their accuracy). These factors also play a role in the level of confidence expressed, and whether witnesses repeat the same decisions across repeated lineups. Despite of these factors, the confidence-accuracy relationship remained intact. This further shows that the confidence-accuracy relationship is robust. In addition to accuracy, confidence is also a predictor of reproducibility. That is, it predicts the likelihood of repeated decisions across repeated lineups. Lastly, confidence judgments are a more reliable indicator of accuracy than response time.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: an archival analysis. *Law and Human Behavior*, *25*(5), 475.

Bothwell, R. K., Deffenbacher, K. a., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, *72*(4), 691–695. https://doi.org/10.1037/0021-9010.72.4.691

Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology*, *87*(1), 112.

Bradfield, A., & McQuiston, D. E. (2004). When does evidence of eyewitness confidence inflation affect judgments in a criminal trial? *Law and Human Behavior*, *28*(4), 369–387. https://doi.org/10.1023/B:LAHU.0000039331.54147.ff

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, *26*(3), 353–364. https://doi.org/10.1023/A:1015380522722

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior*, *30*(1), 31–50. https://doi.org/10.1007/s10979-006-9002-7

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology. Applied*, *12*(1), 11–30. https://doi.org/10.1037/1076-898X.12.1.11

Brigham, J. C., & Cairns, D. L. (1988). The Effect of Mugshot Inspections on Eyewitness Identification Accuracy. *Journal of Applied Social Psychology*, *18*(16), 1394–1410.

Cutler, B. L., Penrod, S. D., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the Effects of Contextual Cues on Eyewitness Identification Accuracy. Social Behaviour.

Deffenbacher, K. A., Bornstein, B. H., Mcgorty, E. K., & Penrod, S. D. (2008). Forgetting the Once-Seen Face : Estimating the Strength of an Eyewitness ' s Memory Representation, *14*(2), 139–150. https://doi.org/10.1037/1076-898X.14.2.139

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, *30*(3), 287–307. https://doi.org/10.1007/s10979-006-9008-1

Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior*, *6*(1), 15–30. https://doi.org/10.1007/BF01049310

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: a 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *The Journal of Applied Psychology*, *87*(5), 951–962. https://doi.org/10.1037/0021-9010.87.5.951

Egan, D., Pittner, M., Goldstein, A. G., Egan, D., Pittner, M., & Goldstein, A. G. (1977). Photographs vs . Live Models, *1*(2), 199–206.

Garrett, B. (2011). *Convicting the innocent*. Harvard University Press.

Godfrey, R. D., & Clark, S. E. (2010). Repeated eyewitness identification procedures: Memory, decision making, and probative value. *Law and Human Behavior*, *34*(3), 241–258.

Hinz, T., & Pezdek, K. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law and Human Behavior*, *25*(2), 185–198. https://doi.org/10.1023/A:1005697431830

Hothorn, T., Bretz, F., Westfall, P., & Heiberger, R. M. (2012). Multcomp: simultaneous inference for general linear hypotheses. *UR L http://CRAN. R-Project. Org/package= Multcomp, R Package Version*, 1–2.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1304–1316. https://doi.org/10.1037/0278-7393.22.5.1304

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80–113. https://doi.org/10.1037/a0025648

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349–358.

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: the effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition*, *30*(8), 1219–1227. https://doi.org/10.3758/BF03213404

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*(4), 361–376. https://doi.org/10.1037/a0030609

Palmer, M. a, Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology. Applied*, *19*(1), 55–71. https://doi.org/10.1037/a0031602

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening line-up most likely to affect eyewitness identification accuracy? *Legal and Criminological Psychology*, *10*(2), 247–263. https://doi.org/10.1348/135532505X49846

Potter, R., & Brewer, N. (1999). Perceptions of witness behaviour-accuracy relationships held by police, lawyers and mock-jurors. *Psychiatry, Psychology and Law*, *6*(1), 97–103. https://doi.org/10.1080/13218719909524952

Quinlivan, D. S., Neuschatz, J. S., Douglass, A. B., Wells, G. L., & Wetmore, S. A. (2012). The effect of post-identification feedback, delay, and suspicion on accurate eyewitnesses. *Law and Human Behavior, 36*(3), 206–214. https://doi.org/10.1037/h0093970

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*(4), 337–347. https://doi.org/10.1007/s10979-009-9192-x

Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *The Journal of Applied Psychology*, *89*(2), 334–346. https://doi.org/10.1037/0021-9010.89.2.334

Shapiro, P. N., & Penrod, S. (1986). Meta-Analysis of Facial Identification Studies, *100*(2), 139–156.

Shepherd, J. W. (1983). Identification after long delays. *Evaluating Witness Evidence*, 173–187.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of Eyewitness Errors : Can False Identifications Be Diagnosed ? *Journal of Applied Psychology*, *85*(4), 542–550. https://doi.org/10.1037//0021-9010.85.4.542

Sporer, S. L., Penrod, S., Cutler, B., & Read, D. (1995). Choosing , Confidence , and Accuracy : A Meta-Analysis of the Confidence-Accuracy Relation in Eyewitness Identification Studies, *118*(3).

Steblay, N. K. (2011). What We Know Now : The Evanston Illinois Field Lineups, 1–12. https://doi.org/10.1007/s10979-009-9207-7

Steblay, N. K., Tix, R. W., & Benson, S. L. (2013). Double Exposure : The Effects of Repeated Identi fi cation Lineups on Eyewitness Accuracy, *654*(August), 644–654.

Tekin, E., & Roediger, H. L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy. *Cognitive Research: Principles and Implications*, 1–13. https://doi.org/10.1186/s41235-017-0086-z

Valentine, T., Davis, J. P., Memon, A., & Roberts, A. (2012). Live Showups and Their Influence on a Subsequent Video Line-up. *Applied Cognitive Psychology*, *26*(1), 1–23. https://doi.org/10.1002/acp.1796

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology*, *17*(8), 969–993. https://doi.org/10.1002/acp.939

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335. https://doi.org/10.1016/j.jml.2003.10.003

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology*, *20*(1), 17–31. https://doi.org/10.1002/acp.1166

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: the unruly 10-12-second rule. *Journal of Experimental Psychology: Applied*, *10*(3), 139–47. https://doi.org/10.1037/1076-898X.10.3.139

Wells, G. L., & Bradfield, A. L. (1998). “ Good , you identified the suspect ”: Feedback to eyewitnesses distorts their reports of the witnessing experience . *Journal of Applied Psychology*, *83*(3). https://doi.org/10.1037/0021-9010.83.3.360

Wells, G. L., Olson, E. A., & Charman, S. D. (2003). Distorted Retrospective Eyewitness Reports as Functions of Feedback and Delay, *9*(1), 42–52. https://doi.org/10.1037/1076-898X.9.1.42

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*, *70*(6), 515.

Wixted, J. T., Read, J. D., & Lindsay, D. S. (2016). The Effect of Retention Interval on the Eyewitness Identification Confidence – Accuracy Relationship. *Journal of Applied Research in Memory and Cognition*, *5*(2), 192–203. https://doi.org/10.1016/j.jarmac.2016.04.006

Wixted, J. T., & Wells, G. L. (2017). The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science in the Public Interest*, *18*(1), 10–65. https://doi.org/10.1177/1529100616686966

Yap, M. J., Balota, D. a, Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(3), 495–513. https://doi.org/10.1037/0278-7393.34.3.495

Appendix A

Experiment 1: Average Response Times

Prior to analysis, outliers in the response time data was trimmed. A hard cutoff of 60 seconds was used because roughly only 2% of participants took longer than 60 seconds. Response times that were faster than 500 ms were also removed because it was too fast for a decision involving an array of six photos. This lower end cutoff only removed one participant from the target-absent lineup data (a RT of 106 ms). Prior to trimming, the skewness and kurtosis values of the target-present data were 7.96 and 89.87, respectively. By applying these hard cutoffs, the skewness and kurtosis values were reduced to 1.99, and 4.46, respectively. Again, the same data trimming procedure was applied to the target-absent lineups. The skewness and kurtosis values for target-absent lineups were initially 14.45 and 275.82. After trimming, they were 2.19 and 5.36, respectively. In both cases, there was a tremendous reduction in skewness and kurtosis. Because the response time data were still positively skewed, they were normalized via a log transformation for analysis purposes, but the raw values were reported. These procedures were applied to both the target-present and the target-absent data.

Similar to the confidence rating data, the response time data was also analyzed with linear mixed-effects regression (LMER, using the lme4 package in R). The LMER results were reported in terms of F tests. Degrees of freedom were approximated with a Type II Satterthwaite approximation. All follow-up tests were analyzed using simultaneous tests for general linear hypotheses via the multcomp package in R.

Target-Present Lineups. Figure A1 shows the average response time of choosers and nonchoosers in target-present Lineup 1 and Lineup 2. We performed a 2 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. Results showed a significant main

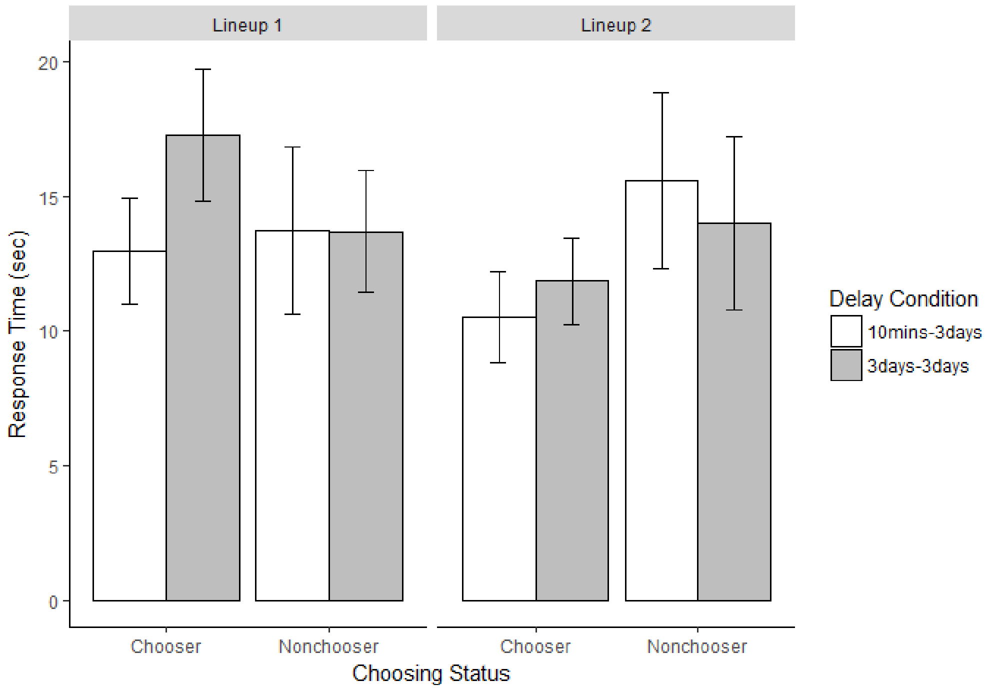
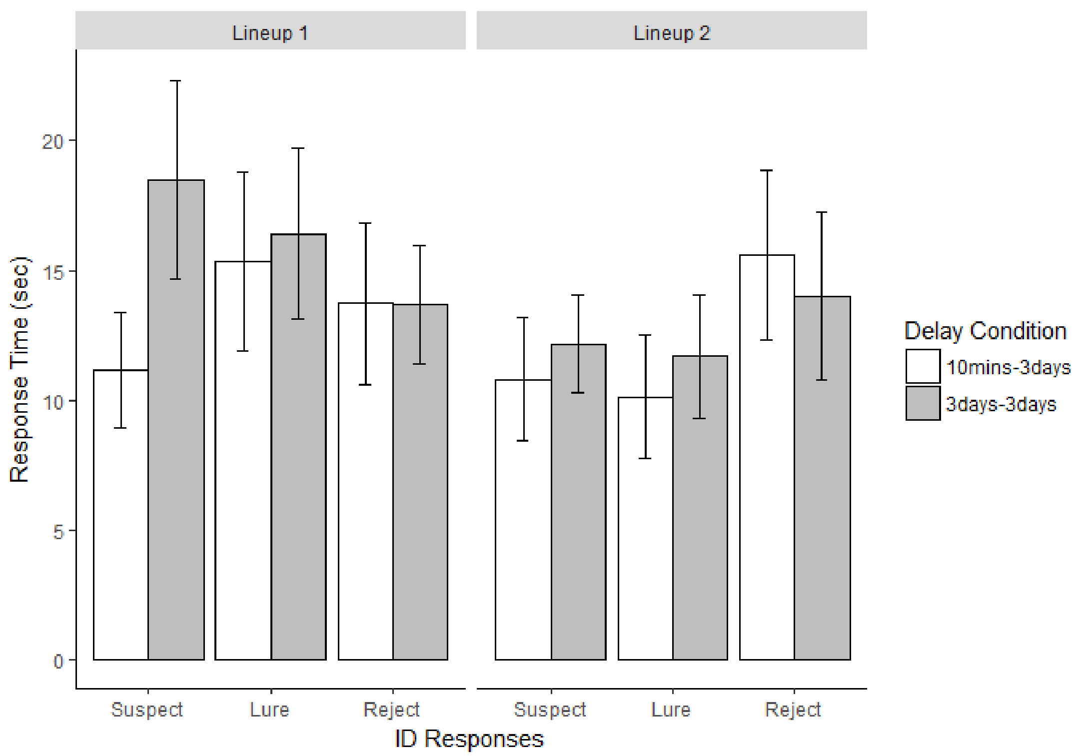


Figure A1. The average response time of choosers and nonchoosers in Experiment 1 target-present Lineup 1 and Lineup 2.



Figures A2. The average response time for all the three response types in Experiment 1 target-present lineups.

effect of delay, $F(1,281.46) = 4.28$, $p < .05$, and a main effect of repeated lineups, $F(1,282.29) = 30.63$, $p < .001$. Participants in the 10-min delay condition made their decisions faster than those in the 3-day condition. Lineup 2 decisions were faster than Lineup 1 decisions. There was no main effect of choosing status. There was a two-way interaction between delay and repeated lineups, $F(1,282.04) = 6.73$, $p < .05$, and a two-way interaction between repeated lineups and choosing status, $F(1,335.40) = 23.55$, $p < .001$. Follow-up tests were conducted for the delay x repeated lineups interaction. In Lineup 1, participants in the 10-min delay condition made their decisions faster than those in 3-day delay condition, $z = -2.31$, $p < .05$; however, they did not differ in Lineup 2. Next, follow-up tests were also conducted for the repeated lineups x choosing status interaction. Choosers were faster in Lineup 2 than Lineup 1, $z = 4.91$, $p < .001$, but nonchoosers did not differ across repeated lineups. In sum, participants in the 10-min delay condition were generally faster than those in the 3-day delay condition; however, this difference was only apparent in Lineup 1. Although participants made their Lineup 2 decisions faster than their Lineup 1 decisions, only choosers became significantly faster in Lineup 2.

Next, we examined whether correct choosers were faster than incorrect choosers. The average response time for choosers were divided into correct (suspect IDs) and incorrect choosers (lure IDs). Figures A2 shows the average response time for all the three response types in target-present lineups. Again, choosers made their Lineup 2 identification decisions faster than their Lineup 1 decisions. We performed a 2 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 3 (identification responses) mixed model. There was a main effects of delay, $F(1,282.26) = 4.57$, $p < .05$, and a main effect of repeated lineups, $F(1,279.20) = 30.05$, $p < .001$. Participants in the 10-min delay condition made their decisions faster than those in the 3-day condition. Lineup 2 decisions were faster than Lineup 1 decisions. There was no main effect of identification responses. There was a two-way interaction between delay and repeated lineups, $F(1,278.59) = 5.53$, $p<.05$, and a two-way interaction between repeated lineups and identification responses, $F(2,319.64) = 12.61$, $p<.001$. Follow-up tests of the interaction between delay and repeated lineups were conducted. In Lineup 1, participants in the 10-min delay condition made their decisions faster than those in 3-day delay condition, $z = -2.96$, $p <.05$; however, they did not differ in Lineup 2. Next, the follow-up tests were conducted for the interaction between repeated lineups and identification responses. In Lineup 1, suspect IDs, lure IDs, and "Not Present" responses did not differ. In Lineup 2, participants made lure IDs faster than "Not Present" responses, $z = -3.30$, $p <.05$. The difference between suspect IDs and "Not Present" responses did not reach significance. It appears that participants made suspect and lure IDs with similar response times. In addition, when a lineup was repeated, they were quick to make a lure ID than a "Not Present" response.

Target-Absent Lineups. Figure A3 shows the average response time for choosers and nonchoosers for target-absent Lineup 1 and Lineup 2. We conducted a 2 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. There was a significant main effect of repeated lineups, $F(1,286.21) = 43.00$, $p <.001$. Lineup 2 decisions were faster than Lineup 1 decisions. No main effect of delay or choosing status. There was a two-way interaction between delay and repeated lineup, $F(286.77) = 3.91$, $p < .05$, and a two-way interaction between repeated lineups and choosing status, $F(1,329.74) = 9.74$, $p<.05$. Follow-up tests were conducted on the delay x repeated lineups interaction. In both Lineup 1 and Lineup 2, participants in the 10-min delay and the 3-day delay condition did not differ in terms of response times. Next, we performed follow-up tests on the repeated lineups x choosing status interaction. Choosers were faster in Lineup 2 than Lineup 1, $z = 3.87$, $p <.001$, but nonchoosers did not differ across

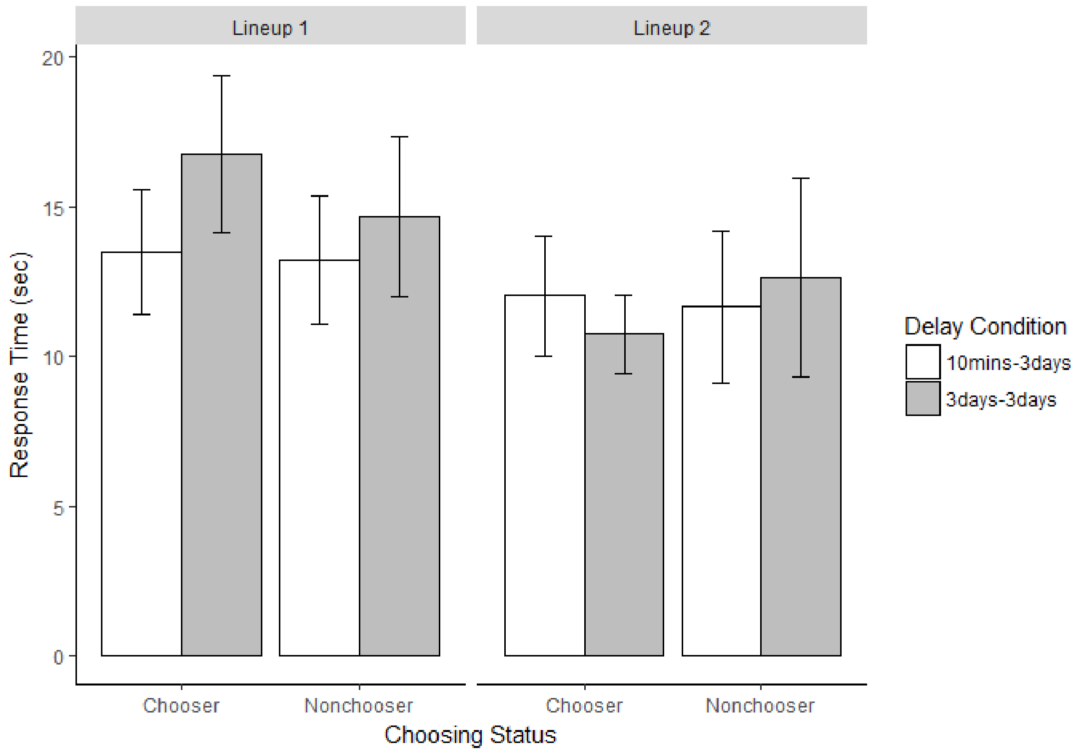


Figure A3. The average response time for choosers and nonchoosers for target-absent Lineup 1 and Lineup 2.

repeated lineups. In other words, choosers were making errors with faster in Lineup 2 than in Lineup 1.

Experiment 2: Average Response Times

Prior to analysis, outliers in the response time data was trimmed. A hard cutoff of 60 seconds was used because roughly only 2% of participants took longer than 60 seconds. Response times that were faster than 500ms were also removed because it was too fast for an array of six photos. This only removed one participant from the target-absent lineup data (a RT of 19 ms). Because the response time data was positively skewed, it normalized via log transformation for analysis purposes, but the raw values were reported in the present study. These procedures were applied to both the target-present and the target-absent data.

Similar to the confidence rating data, the response time data was also analyzed with linear mixed-effects regression (LMER, using the lme4 package in R). The LMER results were reported in terms of F tests. Degrees of freedom were approximated with Type II Satterthwaite approximation. All follow-up tests were analyzed using simultaneous tests for general linear hypotheses via the multcomp package in R.

Target-Present Lineups. Figure A4 shows the average response time of choosers and nonchoosers in target-present Lineup 1 and Lineup 2. The average response time of choosers and nonchoosers was submitted a 4 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. Results indicated significant a main effect of delay, $F(3,426.23) = 8.77$, $p < .001$, a main effect of repeated lineups, $F(1,422.38) = 81.90$, $p < .001$, and a main effect of choosing status, $F(1,855.13) = 5.03$, $p < .05$. Participants in the short-initial delay conditions made their decisions faster than those in the long-initial delay conditions. They also made Lineup 2 decisions faster than their Lineup 1 decisions. Choosers were faster than nonchoosers. There

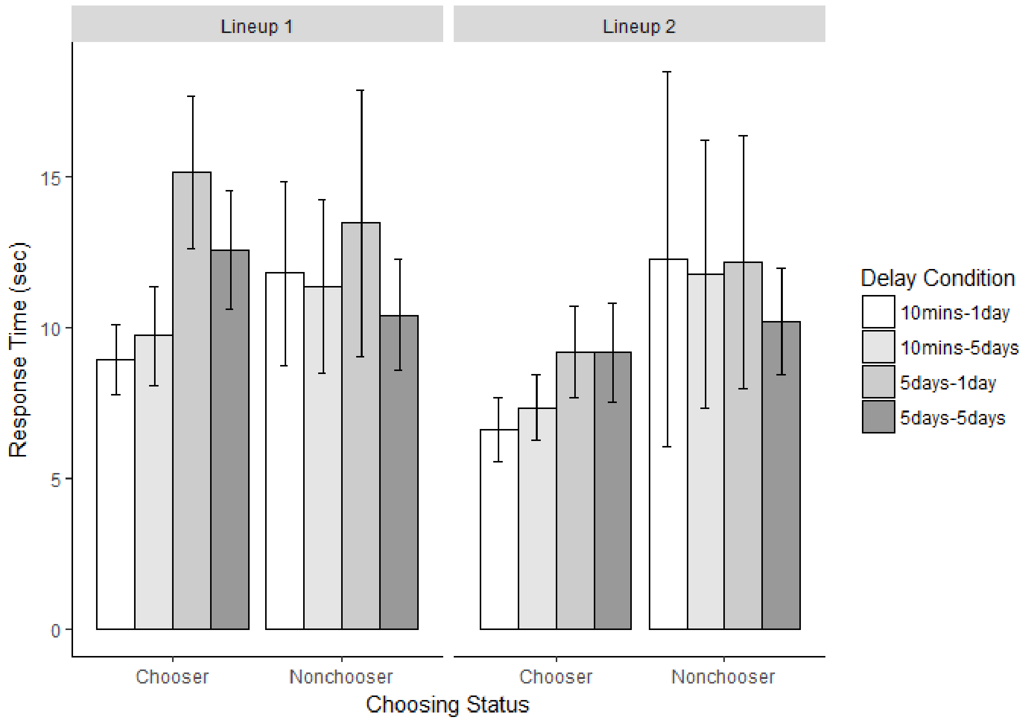


Figure A4. The average response time of choosers and nonchoosers in Experiment 2 target-present Lineup 1 and Lineup 2.

was also a two-way interaction between repeated lineups and choosing status, $F(1,572.72) = 13.05$, $p < .001$. Follow-up tests of the repeated lineups x choosing status interaction revealed that choosers were significantly faster in Lineup 2 than Lineup 1, $z = 7.30$, $p < .001$, but this difference was not observed in nonchoosers.

Figure A5 shows the average response time for all the possible response types in target-present lineups. The average response time of suspect IDs, lure IDs and "Not Present" responses was submitted to a 4 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 3 (identification responses) mixed model. Results indicated significant a main effect of delay, $F(3,426.02) = 7.13$, $p < .001$, a main effect of repeated lineups, $F(1,420.72) = 83.78$, $p < .001$, and a main effect of identification responses, $F(2,785.54) = 6.71$, $p < .05$. Again, participants in the short-initial delay conditions made their decisions faster than those in the long-initial delay conditions. They also made Lineup 2 decisions faster than their Lineup 1 decisions. Suspect IDs were made with faster response times than lure IDs and "Not Present" responses. There was also a significant two-way interaction between repeated lineups and identification responses, $F(2,522.55) = 7.37$, $p < .001$. Follow-up tests were conducted for the repeated lineups x identification responses interaction. For Lineup 2, the average RT of suspect IDs was faster than lure IDs, $z = -2.88$, $p < .05$, and "Not Present" responses, $z = -4.97$, $p < .001$. The average RT of lure IDs was faster than "Not Present" responses, $z = -3.06$, $p < .05$. While these RTs were significantly different across in Lineup 2 identification decisions, they were not for Lineup 1.

Target-Absent Lineups. Figure A6 shows the average RT of choosers and nonchoosers in target-absent Lineup 1 and Lineup 2. The average RT of choosers and nonchoosers was submitted to a 4 (delay conditions) x 2 (Lineup 1 and Lineup 2) x 2 (choosing status) mixed model. Results indicated that there was a main effect of repeated lineups, $F(1,446.29) = 78.22$,

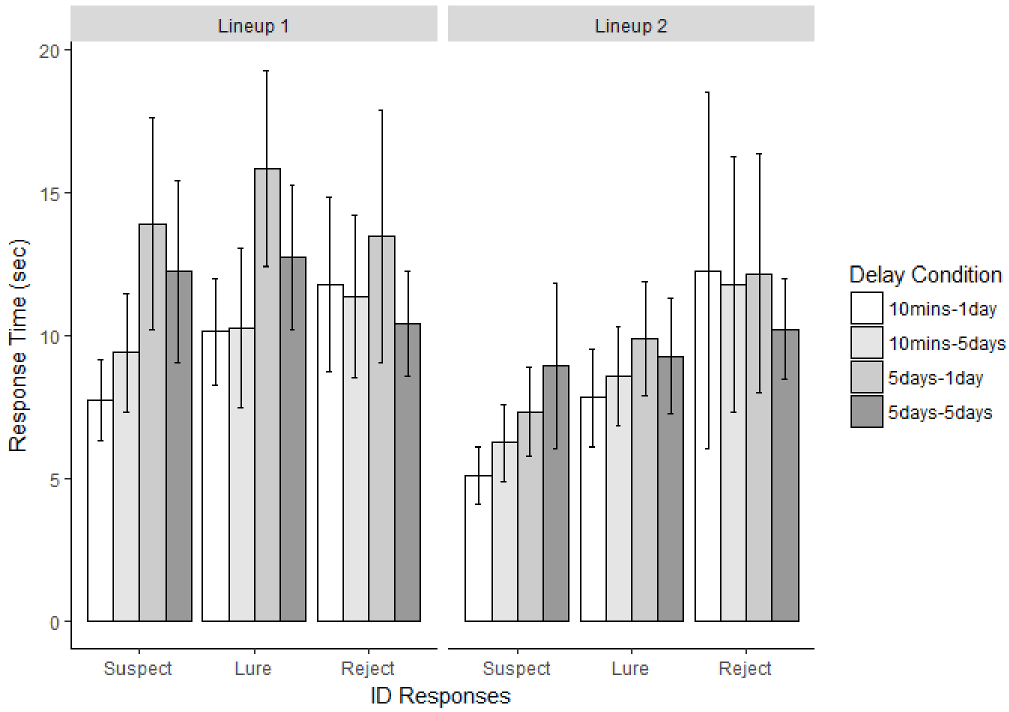


Figure A5. The average response time for all the possible response types in Experiment 2 target-present lineups.

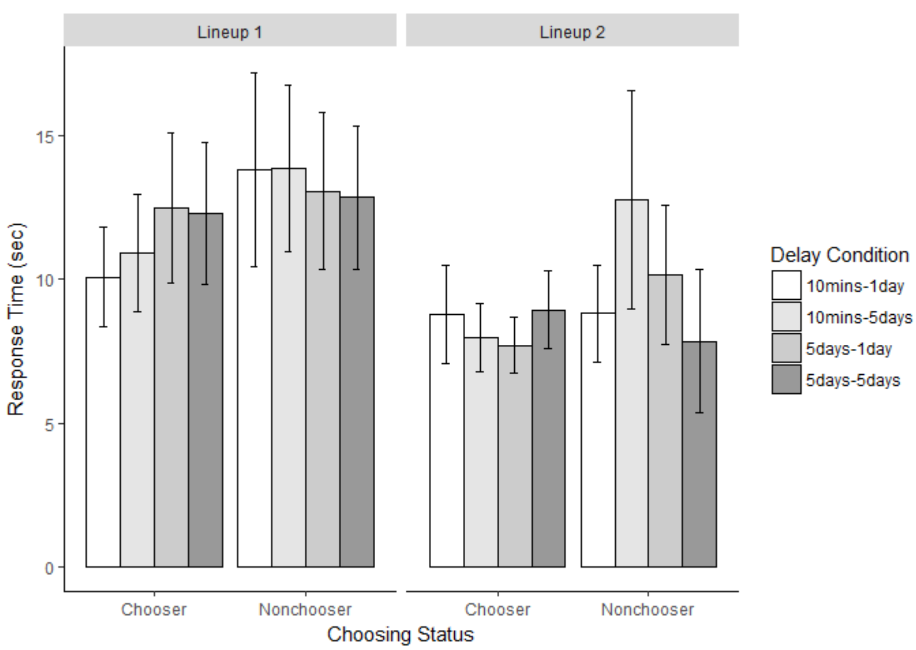


Figure A6. The average RT of choosers and nonchoosers in Experiment 2 target-absent Lineup 1 and Lineup 2.

$p < .001$, and a main effect of choosing status, $F(1,853.93) = 14.23$, $p < .001$. Participants made Lineup 2 decisions faster than their Lineup 1 decisions; choosers were faster than nonchoosers. There was no two-way or three-way interaction. These results showed that choosers were faster than nonchoosers, even though choosers were making an incorrect decision.

Compared to confidence judgments, response times appeared to be a less reliable indicator of accuracy. For target-present lineups, choosers were faster than nonchoosers. Consistent with prior literature, correct choosers tended to be faster than incorrect choosers. However, for target-absent lineups, choosers were also faster than nonchoosers. In other words, false IDs were made faster than correct rejections. One reason this was not observed in Experiment 1 may be explained by difference in the identification procedure between Experiments 1 and 2. In Experiment 1, choosers had to click on a lineup member, then they had to click the "next" button to submit their response, whereas nonchoosers only had to click on "Not Present" to proceed. The "next" button was removed in Experiment 2. Participants were told that only their first choice would be recorded. Whether they decided to choose someone or reject the lineup, they only had to click the lineup member or the "Not Present" button. By equating the number of steps required to make an identification, it appeared that choosers were generally faster than nonchoosers regardless whether they were choosing someone from a target-present or target-absent lineup.

Appendix B

Experiment 1: Optimal Time Boundary

Does response time differentiate correct identifications from incorrect ones? Although correct identification decisions were generally faster than incorrect ones, the response time difference was only a few seconds apart. However, the minor difference may be due to numerous outliers. To better examine whether a specific time boundary best differentiate correct identifications from the incorrect ones, the time-boundary analysis was used (Dunning & Perretta, 2002).

The time-boundary analysis involves separating witnesses who had made a positive identification (i.e. choosers) within 1 s from those who took longer than 1 s. This process is repeated at each time-boundary (1 s, 2 s, 3 s, and so on). For each time boundary, witnesses were further divided into accurate and inaccurate witnesses. Together, this process yields a 2 (accurate vs. inaccurate witnesses) x 2 (inside vs. outside of the specific time boundary) chi-square contingency table for each time boundary. A large chi-square statistic would indicate that the specific time boundary best differentiate correct identifications from the incorrect ones.

Following Dunning and Perretta (2002)'s procedures, the choosers from both target-present and target-absent lineups were combined for the time-boundary analysis. In other words, the correct IDs can only come from the suspect IDs in target-present lineups, whereas incorrect IDs can come from lure IDs in target-present or target-absent lineups.

Consistent with Weber et al., we also found varying boundaries in the present experiment. In the 10-min delay condition, the optimal time boundary occurred roughly around 6 s for Lineup 1 and 7 s for Lineup 2. In the 3-day delay condition, the optimal time boundary occurred roughly around 11 s for Lineup 1 and 13 s for Lineup 2. Given these results, it is difficult to pinpoint a universal optimal time-boundary because the effects of delay and repeated lineups can shift this boundary. That is, for a given eyewitness scenario, we do not have an established guide on determining whether a specific RT is fast or slow, or whether it is diagnostic.

Experiment 2: Optimal Time Boundary

As in Experiment 1, the same time-boundary analysis procedure was conducted for choosers in each of the four delay conditions in the present experiment. Once again, the optimal time boundary for each of delay condition varied. For example, the optimal Lineup 1 time boundaries were 12 s, 8 s, 6 s, and 9 s for the 10mins-1day, 10mins-5days, 5days-1day, and 5days-5days, respectively. These results are consistent with Brewer and Wells (2004)'s findings that the optimal time boundary obtained from time-boundary analysis varied across different manipulations. In the present study, we showed that both delay and repeated lineups also rendered the optimal time boundary unpredictable. In other words, we cannot establish a universal diagnostic RT cutoff.